

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
29 January 2004 (29.01.2004)

PCT

(10) International Publication Number
WO 2004/009838 A2

(51) International Patent Classification⁷: **C12Q**

(21) International Application Number:
PCT/US2003/023057

(22) International Filing Date: 23 July 2003 (23.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/398,156 23 July 2002 (23.07.2002) US
60/424,894 8 November 2002 (08.11.2002) US

(71) Applicant (for all designated States except US): **NEOSE TECHNOLOGIES, INC.** [US/US]; 102 Witmer Road, Horsham, PA 19044 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JOHNSON, Karl, F.** [US/US]; 5320 Ivystream Road, Hatboro, PA 19040 (US).
BEZILA, Daniel, James [US/US]; 715 Red Lion Road, 2nd Floor, Philadelphia, PA 19115 (US).

(74) Agents: **KELLY, Beth, L.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: *H. Pylori* FUCOSYLTRANSFERASES

Linkage Analysis by GC/MS

The samples were methylated, hydrolyzed, reduced with sodium borodeuteride, acetylated and analyzed by GC/MS along with samples of L_NnT and L_NF3.

- A Glc vs. Glc-NAc value close to 1 favors fucosylation of Glc-NAc.
- A Glc vs. Glc-NAc value close to 0 favors fucosylation of Glc

<i>H. Pylori</i> Strain	Glc vs. Glc-NAc
915A2	0.982
19C2A5	0.040
1111A2	0.975
19C2B1	0.991
1182B3	0.983

(57) Abstract: This invention provides nucleic acid and amino acid sequences of fucosyltransferases from *Helicobacter pylori*. The invention also provides methods to use the fucosyltransferases to synthesize oligosaccharides, and glycolipids.



WO 2004/009838 A2

***H. pylori* FUCOSYLTRANSFERASES**

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/398,156, filed July 23, 2002, and U.S. Provisional Application 60/424,894, filed November 8, 2002; both of which are herein incorporated by reference for all purposes.

FIELD OF INVENTION

[0002] This invention provides nucleic acid and amino acid sequences of fucosyltransferases from *Helicobacter pylori*. The invention also provides methods to use the fucosyltransferases to synthesize oligosaccharides, and glycolipids.

BACKGROUND OF THE INVENTION

[0003] Although in recent years significant advances have been made in carbohydrate chemistry, there are still substantial difficulties associated with the chemical synthesis of glycoconjugates, particularly with the formation of the ubiquitous β -1,2-cis-mannoside linkage found in mammalian oligosaccharides. Moreover, regio- and stereo-chemical obstacles must be resolved at each step of the *de novo* synthesis of a carbohydrate.

[0004] In view of the difficulties associated with the chemical synthesis of glycoconjugates, the use of glycosyltransferases to enzymatically synthesize glycoproteins and glycolipids, having desired oligosaccharide moieties, is a promising approach to preparing such glycoconjugates. Enzyme-based syntheses have the advantages of regioselectivity and stereoselectivity, and can be performed using unprotected substrates. Moreover, glycosyltransferases have been used to enzymatically modify oligosaccharide moieties and have been shown to be very effective for producing specific products with good stereochemical and regiochemical control. The glycosyltransferases of interest include fucosyltransferases, sialyltransferases, galactosyltransferases, and N-acetylglucosaminyltransferases. For a general review, see, Crout *et al.*, *Curr. Opin. Chem. Biol.* 2: 98-111 (1998) and Arsequell, *et al.*, *Tetrahedon: Asymmetry* 10: 2839 (1997).

[0005] Many glycoproteins and glycolipids require the presence of a particular glycoform, or the absence of a particular glycoform, in order to exhibit a particular biological activity.

For example, many glycoprotein and glycolipids require the presence of particular

5 fucosylated structures in order to exhibit biological activity. Intercellular recognition mechanisms often require a fucosylated oligosaccharide. For example, a number of glycoproteins that function as cell adhesion molecules, including P-selectin, L-selectin, and E-selectin, bind specific cell surface fucosylated carbohydrate structures such as the sialyl Lewis-x and the sialyl Lewis-a structures. In addition, the specific carbohydrate structures that form the ABO blood group system are fucosylated. The carbohydrate structures in each of the three groups share a $\text{Fuc}\alpha 1,2\text{Gal}\beta 1$ -disaccharide unit. In blood group O structures, this disaccharide is the terminal structure; whereas the blood group A structure is formed by an $\alpha 1,3$ GalNAc transferase that adds a terminal GalNAc residue to the disaccharide; and the blood group B structure is formed by an $\alpha 1,3$ galactosyltransferase that adds a terminal galactose residue.

[0006] The Lewis blood group structures are also fucosylated. For example the Lewis-x and Lewis-a structures are $\text{Gal}\beta 1,4(\text{Fuc}\alpha 1,3)\text{GlcNac}$ and $\text{Gal}\beta 1,3(\text{Fuc}\alpha 1,4)\text{GlcNac}$, respectively. Both these structures can be further sialylated ($\text{NeuAc}\alpha 2,3$ -) to form the corresponding sialylated structures. Other Lewis blood group structures of interest are the Lewis-y and Lewis-b structures which are $\text{Fuc}\alpha 1,2\text{Gal}\beta 1,4(\text{Fuc}\alpha 1,3)\text{GlcNac}\beta$ -OR and $\text{Fuc}\alpha 1,2\text{Gal}\beta 1,3(\text{Fuc}\alpha 1,4)\text{GlcNac}$ -OR, respectively. For a description of the structures of the ABO and Lewis blood group structures and the enzymes involved in their synthesis see, *Essentials of Glycobiology*, Varki *et al.* eds., Chapter 16 (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1999).

[0007] Specifically, fucosyltransferases have been used in synthetic pathways to transfer a fucose residue from guanosine-5'-diphosphofucose to a specific hydroxyl of a saccharide acceptor. A variety of donor substrates and acceptor substrates are known (see Guo *et al.*, *Applied Biochem. and Biotech.* **68**: 1-20 (1997)). For example, Ichikawa prepared sialyl Lewis-x by a method that involves the fucosylation of sialylated lactosamine with a cloned fucosyltransferase (Ichikawa *et al.*, *J. Am. Chem. Soc.* **114**: 9283-9298 (1992)). Lowe has described a method for expressing non-native fucosylation activity in cells, thereby producing fucosylated glycoproteins on cell surfaces, *etc.* (U.S. Patent No. 5,955,347).

[0008] Thus, since the biological activity of many commercially important recombinantly and transgenically produced glycoproteins and glycolipids depends upon the presence of a particular glycoform, or the absence of a particular glycoform, a need exists for an efficient method for enzymatically synthesizing glycoconjugates having the desired fucosylated

oligosaccharide moieties. In addition, there is a need for the efficient production of fucosylated oligosaccharides. The present invention fulfills these and other needs.

BRIEF SUMMARY OF THE INVENTION

5 [0009] The present invention provides α -1,3/4-fucosyltransferase proteins and nucleic acids from *H. pylori*. The α -1,3/4-fucosyltransferase proteins catalyze the transfer of a fucose residue from a donor substrate to an acceptor substrate. In one embodiment, the invention provides α -1,3/4-fucosyltransferase nucleic acids with greater than 90% identity to a nucleotide sequence selected from SEQ ID NO:1, 3, or 7 and that encode α -1,3/4-
10 fucosyltransferase proteins that transfer fucose to GlcNAc residues. In another embodiment, the invention provides α -1,3/4-fucosyltransferase nucleic acids with greater than 90% identity to SEQ ID NO:5 and that encode α -1,3/4-fucosyltransferase proteins that transfer fucose to Glucose residues.

[0010] In another embodiment the α -1,3/4-fucosyltransferase nucleic acid is selected from
15 SEQ ID NO:1, 3, 5 or 7. The invention also provides nucleic acid sequences that encode α -1,3/4-fucosyltransferase proteins, including SEQ ID NO:2, 4, 6, or 8 and that catalyze the transfer of fucose to an N-acetylglucosamine residue or to a glucose residue. In one aspect the encoded α -1,3/4-fucosyltransferase also includes an amino acid tag.

[0011] In a further aspect the invention provides an isolated nucleic acid that includes SEQ
20 ID NO:11, and that encodes an α -1,3/4-fucosyltransferase protein that catalyzes the transfer of a fucose residue from a donor substrate to a glucose residue. In another aspect the invention provides a nucleic acid that encodes SEQ ID NO:12.

[0012] In another embodiment the invention provides expression vectors that include the above described α -1,3/4-fucosyltransferase nucleic acids, host cells that include the expression
25 vectors, and methods to produce the α -1,3/4-fucosyltransferase proteins using the host cells cultured under conditions suitable for expression of the α -1,3/4-fucosyltransferase protein.

[0013] In another embodiment the invention provides recombinant fucosyltransferase proteins that include amino acid sequence having greater than 90% identity to SEQ ID NO:2, 4, or 8, wherein the fucosyltransferase catalyzes the transfer of a fucose residue from a donor
30 substrate to N-acetylglucosamine. In another embodiment the invention provides recombinant fucosyltransferase proteins that include amino acid sequence having greater than 90% identity to SEQ ID NO:6, wherein the fucosyltransferase catalyzes the transfer of a

fucose residue from a donor substrate to glucose. In one aspect, the fucosyltransferase proteins comprise SEQ ID NO:2, 4, 6, or 8. In another aspect the fucosyltransferase proteins also include an amino acid tag.

[0014] In another embodiment the invention provides recombinant fucosyltransferase proteins that include SEQ ID NO:12, and that catalyzes the transfer of a fucose residue from a donor substrate to glucose. In another aspect the fucosyltransferase proteins also include an amino acid tag.

[0015] The present invention also provides methods to use the above α -1,3/4-fucosyltransferase protein to produce fucosylated oligosaccharides. The fucosylated oligosaccharides can be further purified. The acceptor substrate can be either N-acetylglucosamine or glucose depending on the needs of the user. In one embodiment the acceptor substrate is Lacto-N-neo-Tetraose (LNnT) and the fucosylated product is Lacto-N-Fucopentaose III (LNFP III). The α -1,3/4-fucosyltransferase can be used in combination with other glycosyltransferases to produce a fucosylated oligosaccharide. For example, using lactose as a starting material, LNFP can be produced through the action of an α -1,3/4-fucosyltransferase that transfers fucose to N-acetylglucosamine, a β -1,3-N-acetylglucosaminyltransferase, and a β -1,4-galactosyltransferase. The β -1,3-N-acetylglucosaminyltransferase and the β -1,4-galactosyltransferase can be bacterial enzymes and in a preferred embodiment are from *Neisseria gonococcus*.

[0016] In another embodiment, the α -1,3/4-fucosyltransferase protein of the present invention are used to produce fucosylated glycolipids. The acceptor substrate can be either N-acetylglucosamine or glucose depending on the needs of the user.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] Figure 1 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 1182B.

[0018] Figure 2 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 1111A.

[0019] Figure 3 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 1218B.

[0020] Figure 4 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 19C2B.

[0021] Figure 5 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 915A.

[0022] Figure 6 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 26695A.

5 [0023] Figure 7 provides the nucleic acid and amino acid sequences of fucosyltransferase from *H. pylori* strain 19C2A.

[0024] Figure 8 provides an alignment between 1182 futB amino acid sequence and a consensus sequence from the glycosyltransferase family 10, *i.e.*, the fucosyltransferase family. Amino acids 23 through 305 of 1182 futB are shown in the top line and represent the
10 most conserved region of the protein, *i.e.* the fucosyltransferase catalytic domain.

[0025] Figure 9 provides an alignment between 1111 futA amino acid sequence and a consensus sequence from the glycosyltransferase family 10, *i.e.*, the fucosyltransferase family. Amino acids 27 through 417 of 1182 futB are shown in the top line and represent the most conserved region of the protein, *i.e.* the fucosyltransferase catalytic domain.

15 [0026] Figure 10 provides an alignment between 1218 futB amino acid sequence and a consensus sequence from the glycosyltransferase family 10, *i.e.*, the fucosyltransferase family. Amino acids 23 through 399 of 1182 futB are shown in the top line and represent the most conserved region of the protein, *i.e.* the fucosyltransferase catalytic domain.

[0027] Figure 11 provides an alignment between 19C2 futB amino acid sequence and a
20 consensus sequence from the glycosyltransferase family 10, *i.e.*, the fucosyltransferase family. Amino acids 23 through 377 of 1182 futB are shown in the top line and represent the most conserved region of the disclosed protein, *i.e.* the fucosyltransferase catalytic domain.

[0028] Figure 12 provides an alignment between amino acid sequence of *H. pylori* strains 1182 FutB, 1111 FutA, 1218 FutB, 19C2 FutB, 915FutA, 19C2 FutA, and 26695 FutA. The
25 bottom sequence is a consensus sequence.

[0029] Figure 13 provides an alignment between nucleic acid sequence of *H. pylori* strains 1182 FutB, 1111 FutA, 1218 FutB, 19C2 FutB, 915FutA, 19C2 FutA, and 26695 FutA. The bottom sequence is a consensus sequence.

[0030] Figure 14 provides oligosaccharide structures of Lacto-N-neo-Tetraose (LNnT), a substrate of the *H. pylori* fucosyltransferases and Lacto-N-Fucopentaose III (LNFPIII or LNFIII), a product of the *H. pylori* fucosyltransferases.

[0031] Figure 15 provides the results of analysis of acceptor specificity for the *H. pylori* fucosyltransferases.

[0032] Figure 16 provides the yield of LNFIII synthesis using the *H. pylori* fucosyltransferases. Two ion exchange resins were tested: MR3 NH_4HCO_3 and Dowex1/Dowex50 resin.

DEFINITIONS

[0033] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell culture, molecular genetics, organic chemistry and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for nucleic acid and peptide synthesis. Generally, enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (*see generally*, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference), which are provided throughout this document. The nomenclature used herein and the laboratory procedures in analytical chemistry, and organic synthetic described below are those well known and commonly employed in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses.

[0034] The terms " α -1,3/4-fucosyltransferase or fucosyltransferase" or a nucleic acid encoding an " α -1,3/4-fucosyltransferase or fucosyltransferase" refer to nucleic acid and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have an amino acid sequence that has greater than about 60% amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of at least about 25, 50, 100, 200, 500, 1000, or more amino acids, to a polypeptide encoded by a nucleic acid selected

from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:13; or an amino acid sequence of SEQ ID NO:2; , SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:14; (2) specifically bind to antibodies, e.g., polyclonal antibodies, raised against an immunogen comprising an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:14; immunogenic fragments thereof, and conservatively modified variants thereof; (3) specifically hybridize under stringent hybridization conditions to a nucleic acid encoding SEQ ID NO:2; , SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:14; e.g., a nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:13; or its complement, and conservatively modified variants thereof; (4) have a nucleic acid sequence that has greater than about 90%, preferably greater than about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity, preferably over a region of at least about 25, 50, 100, 200, 500, 1000, or more nucleotides, to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:13; or its complement. The nucleic acids and proteins of the invention include both naturally occurring or recombinant molecules.

[0035] The α -1,3/4-fucosyltransferase enzymes of the invention can also be recognized by the presence of highly conserved catalytic domains that are found in a family of fucosyltransferase proteins, glycosyltransferase family 10, see *e.g.*, gnl|CDD|16836 pfam00852, Glyco_transf_10. Alignments between conserved catalytic domains of 1182 futB, 1111 futA, 1218 futB, and 19C2 futB and a consensus sequence from the catalytic domain of glycosyltransferase family 10 members are shown in figures 8-11.

[0036] A biologically active fucosyltransferase as described herein is a fucosyltransferase that catalyzes the transfer of fucose from a donor substrate, for example, GDP-fucose, to an acceptor molecule in an α -1,3/4-linkage. The acceptor molecule can be either N-acetylglucosylamine (GlcNAc) or glucose. For example, Fucosyltransferases from the following *H. pylori* strains transfer fucose to Glc-NAc: Strain 915 FutA, Strain 1111 FutA, Strain 19C2 FutB, and Strain 1182 FutB. The FutA gene product from *H. pylori* Strain 19C2 FutA transfers fucose to the reducing glucose of the LNnT acceptor, as did the FutB gene product from *H. pylori* strain 1218, and a novel 26695 FutA protein. In preferred embodiments, the fucosyltransferase transfers fucose exclusively to GlcNAc or exclusively to glucose. The acceptor molecule can be a carbohydrate, an oligosaccharide, or a glycolipid.

[0037] The *H. pylori* fucosyltransferase proteins of the invention are useful for transferring a saccharide from a donor substrate to an acceptor substrate. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. However, in some embodiments the fucose residue is added to a reducing glucose residue.

5 Biomolecules as defined here include but are not limited to biologically significant molecules such as carbohydrates, oligosacchrides, proteins (*e.g.*, glycoproteins), and lipids (*e.g.*, glycolipids, phospholipids, sphingolipids and gangliosides).

[0038] The following abbreviations are used herein:

Ara = arabinosyl;

10 Fru = fructosyl;

Fuc = fucosyl;

Gal = galactosyl;

GalNAc = N-acetylgalactosylamino;

Glc = glucosyl;

15 GlcNAc = N-acetylglucosylamino;

Man = mannosyl; and

NeuAc = sialyl (N-acetylneuraminy)

FT or Fut = fucosyltransferase*

ST = sialyltransferase*

20 GalT = galactosyltransferase*

[0039] Oligosaccharides are considered to have a reducing end and a non-reducing end, whether or not the saccharide at the reducing end is in fact a reducing sugar. In accordance with accepted nomenclature, oligosaccharides are depicted herein with the non-reducing end on the left and the reducing end on the right.

[0040] All oligosaccharides described herein are described with the name or abbreviation for the non-reducing saccharide (*e.g.*, Gal), followed by the configuration of the glycosidic bond (α or β), the ring bond, the ring position of the reducing saccharide involved in the bond, and then the name or abbreviation of the reducing saccharide (*e.g.*, GlcNAc). The linkage between two sugars may be expressed, for example, as 2,3, 2 \rightarrow 3, or (2,3). Each
30 saccharide is a pyranose or furanose.

[0041] The term "sialic acid" refers to any member of a family of nine-carbon carboxylated sugars. The most common member of the sialic acid family is N-acetyl-neuraminic acid (2-keto-5-acetamido-3,5-dideoxy-D-glycero-D-galactononulopyranos-1-onic acid (often abbreviated as Neu5Ac, NeuAc, or NANA). A second member of the family is N-glycolyl-neuraminic acid (Neu5Gc or NeuGc), in which the N-acetyl group of NeuAc is hydroxylated. A third sialic acid family member is 2-keto-3-deoxy-nonulosonic acid (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* **261**: 11550-11557; Kanamori *et al.*, *J. Biol. Chem.* **265**: 21811-21819 (1990)). Also included are 9-substituted sialic acids such as a 9-O-C₁-C₆ acyl-Neu5Ac like 9-O-lactyl-Neu5Ac or 9-O-acetyl-Neu5Ac, 9-deoxy-9-fluoro-Neu5Ac and 9-azido-9-deoxy-Neu5Ac. For review of the sialic acid family, *see, e.g.*, Varki, *Glycobiology* **2**: 25-40 (1992); *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, New York (1992)). The synthesis and use of sialic acid compounds in a sialylation procedure is disclosed in international application WO 92/16640, published October 1, 1992.

[0042] An "acceptor substrate" for a glycosyltransferase is an oligosaccharide moiety that can act as an acceptor for a particular glycosyltransferase. When the acceptor substrate is contacted with the corresponding glycosyltransferase and sugar donor substrate, and other necessary reaction mixture components, and the reaction mixture is incubated for a sufficient period of time, the glycosyltransferase transfers sugar residues from the sugar donor substrate to the acceptor substrate. The acceptor substrate will often vary for different types of a particular glycosyltransferase.

[0043] An "acceptor substrate" for an *H. pylori* fucosyltransferase is an oligosaccharide moiety that can act as an acceptor for a the *H. pylori* fucosyltransferase. When the acceptor substrate is contacted with the *H. pylori* fucosyltransferase and sugar donor substrate (*e.g.*, GDP-fucose), and other necessary reaction mixture components, and the reaction mixture is incubated for a sufficient period of time, the *H. pylori* fucosyltransferase transfers fucose residues from the GDP-fucose to the acceptor substrate. The acceptor substrate will often vary for different types of a particular fucosyltransferases. For example, the acceptor substrate for a mammalian galactoside 2-L-fucosyltransferase (α 1,2-fucosyltransferase) will include a Gal β 1,4-GlcNAc-R at a non-reducing terminus of an oligosaccharide; this fucosyltransferase attaches a fucose residue to the Gal via an α 1,2 linkage. Terminal Gal β 1,4-GlcNAc-R and Gal β 1,3-GlcNAc-R and sialylated analogs thereof are acceptor substrates for α 1,3 and α 1,4-fucosyltransferases, respectively. These enzymes, however,

attach the fucose residue to the GlcNAc residue of the acceptor substrate. Accordingly, the term “acceptor substrate” is taken in context with the particular glycosyltransferase of interest for a particular application. The *H. pylori* fucosyltransferase described herein will transfer fucose to sialylated or unsialylated acceptor substrates. Some *H. pylori* fucosyltransferase described herein will transfer fucose to glucose residues.

[0044] A “donor substrate” for glycosyltransferases is an activated nucleotide sugar. Such activated sugars generally consist of uridine, guanosine, and cytidine monophosphate derivatives of the sugars (UMP, GMP and CMP, respectively) or diphosphate derivatives of the sugars (UDP, GDP and CDP, respectively) in which the nucleoside monophosphate or diphosphate serves as a leaving group. For example, a donor substrate for fucosyltransferases is GDP-fucose. Donor substrates for sialyltransferases, for example, are activated sugar nucleotides comprising the desired sialic acid. For instance, in the case of NeuAc, the activated sugar is CMP-NeuAc.

[0045] A “substantially uniform glycoform” or a “substantially uniform glycosylation pattern,” when referring to a glycoprotein species, refers to the percentage of acceptor substrates that are glycosylated by the glycosyltransferase of interest (*e.g.*, fucosyltransferase). For example, in the case of the α 1,3 or α 1,4 fucosyltransferase noted above, a substantially uniform fucosylation pattern exists if substantially all (as defined below) of the Gal β 1,4-GlcNAc-R and sialylated or unsialylated analogues thereof are fucosylated in a composition comprising the glycoprotein of interest. It will be understood by one of skill in the art, that the starting material may contain glycosylated acceptor substrates (*e.g.*, fucosylated Gal β 1,4-GlcNAc-R substrates). Thus, the calculated amount of glycosylation will include acceptor substrates that are glycosylated by the methods of the invention, as well as those acceptor substrates already glycosylated in the starting material.

[0046] The term “substantially” in the above definitions of “substantially uniform” generally means at least about 60%, at least about 70%, at least about 80%, or more preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the acceptor substrates for a particular glycosyltransferase are glycosylated (*e.g.*, fucosylated Gal β 1,4-GlcNAc-R substrates).

[0047] The term “substantially identical fucosylation pattern,” refers to a glycosylation pattern of a glycoprotein produced by a method of the invention which is at least about 80%, more preferably at least about 90%, even more preferably at least about 91%, 92%, 93%,

94%, or 95% and still more preferably at least about 96%, 97%, 98% or 99% identical to the fucosylation of a known glycoprotein. "Known fucosylation pattern," refers to a fucosylation pattern of a known glycoprotein from any source having any known level of fucosylation.

[0048] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0049] "Protein", "polypeptide", or "peptide" refer to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a polypeptide. When the amino acids are α -amino acids, either the L-optical isomer or the D-optical isomer can be used. Additionally, unnatural amino acids, for example, β -alanine, phenylglycine and homoarginine are also included. Amino acids that are not gene-encoded may also be used in the present invention. Furthermore, amino acids that have been modified to include reactive groups may also be used in the invention. All of the amino acids used in the present invention may be either the D - or L -isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see*, Spatola, A. F., in CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983).

[0050] The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous

to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques. A "recombinant protein" is one which has been produced by a recombinant cell.

[0051] A "fusion protein" refers to an *H. pylori* fucosyltransferase protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof.

[0052] Components of fusion proteins include "accessory enzymes" and/or "purification or amino acid tags." An "accessory enzyme" as referred to herein, is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate for a fucosyltransferase. An accessory enzyme can, for example, catalyze the formation of a nucleotide sugar that is used as a donor moiety by a fucosyltransferase, *e.g.*, GDP-fucose. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar, *e.g.*, fucose. The recombinant fusion protein of the invention can be constructed and expressed as a fusion protein with a molecular "purification tag" at one end, which facilitates purification of the protein. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include "epitope tags," which are a protein sequence that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A "FLAG tag" is a commonly used epitope tag, specifically recognized by a monoclonal anti-FLAG antibody, consisting of the sequence AspTyrLysAspAspAsp AspLys or a substantially identical variant thereof. Other suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide, which will bind to metal ions such as nickel or cobalt ions. Purification tags also include maltose binding domains and starch binding domains. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacylodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0053] The term “functional domain” with reference to glycosyltransferases, refers to a domain of the glycosyltransferase that confers or modulates an activity of the enzyme, *e.g.*, acceptor substrate specificity, catalytic activity, binding affinity, or other biological or biochemical activity. Examples of functional domains of glycosyltransferases include, but are not limited to, the catalytic domain.

[0054] The terms “expression level” or “level of expression” with reference to a protein refers to the amount of a protein produced by a cell. The amount of protein produced by a cell can be measured by the assays and activity units described herein or known to one skilled in the art. One skilled in the art would know how to measure and describe the amount of protein produced by a cell using a variety of assays and units, respectively. Thus, the quantitation and quantitative description of the level of expression of a protein, *e.g.*, an *H. pylori* fucosyltransferase, can be assayed measuring the enzymatic activity or the units used to describe the activity, or the amount of protein. The amount of protein produced by a cell can be determined by standard known assays, for example, the protein assay by Bradford (1976), the bicinchoninic acid protein assay kit from Pierce (Rockford, Illinois), or as described in U.S. Patent No. 5,641,668.

[0055] The term “enzymatic activity” refers to an activity of an enzyme and may be measured by the assays and units described herein or known to one skilled in the art.

[0056] The term “specific activity” as used herein refers to the catalytic activity of an enzyme, *e.g.*, an *H. pylori* fucosyltransferase protein of the present invention, and may be expressed in activity units. As used herein, one activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5.

[0057] A “catalytic domain” refers to a protein domain, or a subsequence thereof, that catalyzes an enzymatic reaction performed by the enzyme. For example, a catalytic domain of a fucosyltransferase will include a subsequence of the fucosyltransferase sufficient to transfer a fucose residue from a donor to an acceptor saccharide. A catalytic domain can include an entire enzyme, a subsequence thereof, or can include additional amino acid sequences that are not attached to the enzyme, or a subsequence thereof, as found in nature. The α -1,3/4-fucosyltransferase enzymes of the invention can also be recognized by the

presence of highly conserved catalytic domains that are found in a family of fucosyltransferase proteins, glycosyltransferase family 10, see *e.g.*, gnl|CDD|16836 pfam00852, Glyco_transf_10. Alignments between conserved catalytic domains of 1182 futB, 1111 futA, 1218 futB, and 19C2 futB and a consensus sequence from the catalytic domain of glycosyltransferase family 10 members are shown in figures 8-11. Highly conserved regions, similar to a region of the glycosyltransferase family 10 catalytic domain consensus sequence starting at about amino acid 11 and ending at amino acid 301, are found in each of the *H. pylori* α -1,3/4-fucosyltransferase enzymes listed above, *e.g.*, 1182 futB, amino acids 23-305; 1111 futA, amino acids 27-304; 1218 futB, amino acids 23-305; and 19C2 futB amino acids 22-277, and are believed to be the catalytic domains of the enzyme. Thus, polypeptides comprising the above-identified fucosyltransferase catalytic domains can be used in the methods of the invention, *e.g.*, fucosylating oligosaccharides or glycolipids. Nucleic acids that encode the above-identified fucosyltransferase catalytic domains can also be used in the methods of the invention, *e.g.*, production of fucosyltransferase proteins for fucosylating oligosaccharides or glycolipids.

[0058] A “subsequence” refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (*e.g.*, protein) respectively.

[0059] The term “nucleic acid” refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

[0060] A “recombinant expression cassette” or simply an “expression cassette” is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of affecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to be transcribed (*e.g.*, a nucleic acid encoding a desired polypeptide), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription

termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette.

[0061] A “heterologous sequence” or a “heterologous nucleic acid”, as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous glycoprotein gene in a eukaryotic host cell includes a glycoprotein-encoding gene that is endogenous to the particular host cell that has been modified. Modification of the heterologous sequence may occur, *e.g.*, by treating the DNA with a restriction enzyme to generate a DNA fragment that is capable of being operably linked to the promoter. Techniques such as site-directed mutagenesis are also useful for modifying a heterologous sequence.

[0062] The term “isolated” refers to material that is substantially or essentially free from components which interfere with the activity of an enzyme. For a saccharide, protein, or nucleic acid of the invention, the term “isolated” refers to material that is substantially or essentially free from components which normally accompany the material as found in its native state. Typically, an isolated saccharide, protein, or nucleic acid of the invention is at least about 80% pure, usually at least about 90%, and preferably at least about 95% pure as measured by band intensity on a silver stained gel or other method for determining purity. Purity or homogeneity can be indicated by a number of means well known in the art. For example, a protein or nucleic acid in a sample can be resolved by polyacrylamide gel electrophoresis, and then the protein or nucleic acid can be visualized by staining. For certain purposes high resolution of the protein or nucleic acid may be desirable and HPLC or a similar means for purification, for example, may be utilized.

[0063] The term “operably linked” refers to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence.

[0064] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or protein sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

[0065] The phrase “substantially identical,” in the context of two nucleic acids or proteins, refers to two or more sequences or subsequences that have greater than about 60% nucleic acid or amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

[0066] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0067] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally, Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)).

[0068] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschuel *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring

sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for
5 initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to
10 calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN
15 program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

20 [0069] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid
25 sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0070] A further indication that two nucleic acid sequences or proteins are substantially
30 identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with the protein encoded by the second nucleic acid, as described below. Thus, a protein is typically substantially identical to a second protein, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences

are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

[0071] The phrase “hybridizing specifically to” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA.

[0072] The term “stringent conditions” refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 15°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at T_m , 50% of the probes are occupied at equilibrium).

Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is typically at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42° C, or, 5x SSC, 1% SDS, incubating at 65° C, with wash in 0.2x SSC, and 0.1% SDS at 65° C. For PCR, a temperature of about 36° C is typical for low stringency amplification, although annealing temperatures may vary between about 32-48° C depending on primer length. For high stringency PCR amplification, a temperature of about 62° C is typical, although high stringency annealing temperatures can range from about 50° C to about 65° C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90-95° C for 30-120 sec, an annealing phase lasting 30-120 sec, and an extension phase of about 72° C for 1-2 min. Protocols and guidelines for low and high stringency amplification reactions are available,

e.g., in Innis, et al. (1990) *PCR Protocols: A Guide to Methods and Applications* Academic Press, N.Y.

[0073] The phrases “specifically binds to a protein” or “specifically immunoreactive with”, when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

[0074] “Conservatively modified variations” of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are “silent variations,” which are one species of “conservatively modified variations.” Every polynucleotide sequence described herein which encodes a protein also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and UGG which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each “silent variation” of a nucleic acid which encodes a protein is implicit in each described sequence.

[0075] Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids

(typically less than 5%, more typically less than 1%) in an encoded sequence are “conservatively modified variations” where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

5 [0076] One of skill will appreciate that many conservative variations of the fusion proteins and nucleic acid which encode the fusion proteins yield essentially identical products. For example, due to the degeneracy of the genetic code, “silent substitutions” (*i.e.*, substitutions of a nucleic acid sequence which do not result in an alteration in an encoded protein) are an implied feature of every nucleic acid sequence which encodes an amino acid. As described
10 herein, sequences are preferably optimized for expression in a particular host cell used to produce the chimeric glycosyltransferases (*e.g.*, yeast, human, and the like). Similarly, “conservative amino acid substitutions,” in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties (*see*, the definitions section, *supra*), are also readily identified as being highly similar to a particular
15 amino acid sequence, or to a particular nucleic acid sequence which encodes an amino acid. Such conservatively substituted variations of any particular sequence are a feature of the present invention. *See also*, Creighton (1984) *Proteins*, W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also
20 “conservatively modified variations”.

[0077] The practice of this invention can involve the construction of recombinant nucleic acids and the expression of genes in transfected host cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids such as expression vectors
25 are well known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques*, *Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates,
30 Inc. and John Wiley & Sons, Inc., (1999 Supplement) (Ausubel). Suitable host cells for expression of the recombinant *H. pylori* fucosyltransferases are known to those of skill in the art, and include, for example, bacterial cells, including *E. coli*. Eucaryotic cells can also be

used in the present invention, for example insect cells such as Sf 9 cell and yeast or fungal cells (*e.g.*, *Aspergillus niger* or yeast).

[0078] Examples of protocols sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.* (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874; Lomell *et al.* (1989) *J. Clin. Chem.* 35: 1826; Landegren *et al.* (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

DETAILED DESCRIPTION OF THE INVENTION

[0079] The present invention provides novel α -1,3/4-fucosyltransferases, *i.e.*, *H. pylori* fucosyltransferases. In addition, the fucosyltransferases are useful for producing fucosylated oligosaccharides and glycolipids.

[0080] Specifically, α -1,3/4-fucosyltransferases from the following *H. pylori* strains were cloned and analyzed: 915A2, 1111A2, 19C2B1, 1182B3, 19C2A5, 26695, and 1218. Fucosyltransferases from the following *H. pylori* strains transferred fucose to Glc-NAc: 915A2, 1111A2, 19C2FutB, and 1182B3. The FutA gene product from *H. pylori* strain 19C2A5 transferred fucose to the reducing glucose of the L_NT acceptor, as did the FutB gene product from *H. pylori* strain 1218. The ability of FutA gene product from *H. pylori* strain 26695 to transfer fucose to glucose was confirmed.

[0081] A major advantage of the *H. pylori* α -1,3/4-fucosyltransferases over mammalian α -1,3/4-fucosyltransferases is that the *H. pylori* enzyme appears to be unaffected by the sialylation status of the acceptor. In addition some of the *H. pylori* fucosyltransferases add fucose exclusively to the N-acetylglucosamine (glcNAc) residue in acceptor sugars that contain both glucose and glcNAc residues. In contrast, mammalian α -1,3/4-fucosyltransferases are sensitive to the degree of sialylation of the acceptor and some mammalian enzymes add to both glucose and glcNAc residues in the same acceptor. In

addition bacterially expressed enzymes offer a large cost savings relative to the expression of mammalian gene products in Sf9 or CHO systems.

A. Cloning Of *H. pylori* α -1,3/4-fucosyltransferases proteins

5 [0082] Nucleic acids that encode glycosyltransferases, *e.g.*, *H. pylori* α -1,3/4-fucosyltransferases and methods of obtaining such nucleic acids, are known to those of skill in the art. Suitable nucleic acids (*e.g.*, cDNA, genomic, or subsequences (probes)) can be cloned, or amplified by *in vitro* methods such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), or the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) 10 *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook *et al.*); *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994 Supplement) (Ausubel); Cashion *et al.*, U.S. patent number 5,017,478; and Carr, European Patent No. 15 0,246,864.

[0083] A DNA that encodes an *H. pylori* α -1,3/4-fucosyltransferase, or a subsequences thereof, can be prepared by any suitable method described above, including, for example, cloning and restriction of appropriate sequences with restriction enzymes. In one preferred embodiment, nucleic acids encoding *H. pylori* α -1,3/4-fucosyltransferases are isolated by 25 routine cloning methods. A nucleotide sequence of a *H. pylori* α -1,3/4-fucosyltransferase as provided in, for example, GenBank or other sequence database (see above) can be used to provide probes that specifically hybridize to a *H. pylori* α -1,3/4-fucosyltransferases gene in a genomic DNA sample, or to an mRNA, encoding an *H. pylori* α -1,3/4-fucosyltransferase, in a total RNA sample (*e.g.*, in a Southern or Northern blot). Once the target nucleic acid 30 encoding a *H. pylori* α -1,3/4-fucosyltransferase is identified, it can be isolated according to standard methods known to those of skill in the art (*see, e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed., Vols. 1-3*, Cold Spring Harbor Laboratory; Berger and Kimmel (1987) *Methods in Enzymology, Vol. 152: Guide to*

Molecular Cloning Techniques, San Diego: Academic Press, Inc.; or Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York). Further, the isolated nucleic acids can be cleaved with restriction enzymes to create nucleic acids encoding the full-length *H. pylori* α -1,3/4-fucosyltransferase, or subsequences thereof, *e.g.*, containing subsequences encoding at least a subsequence of a catalytic domain of a *H. pylori* α -1,3/4-fucosyltransferase. These restriction enzyme fragments, encoding an *H. pylori* α -1,3/4-fucosyltransferase or subsequences thereof, may then be ligated, for example, to produce a nucleic acid encoding an *H. pylori* α -1,3/4-fucosyltransferase protein.

[0084] A nucleic acid encoding an *H. pylori* α -1,3/4-fucosyltransferase, or a subsequence thereof, can be characterized by assaying for the expressed product. Assays based on the detection of the physical, chemical, or immunological properties of the expressed protein can be used. For example, one can identify a cloned *H. pylori* α -1,3/4-fucosyltransferases, by the ability of a protein encoded by the nucleic acid to catalyze the transfer of a fucose residue from a donor substrate to an acceptor substrate. In one method, capillary electrophoresis is employed to detect the reaction products. This highly sensitive assay involves using either saccharide or disaccharide aminophenyl derivatives which are labeled with fluorescein as described in Wakarchuk *et al.* (1996) *J. Biol. Chem.* 271 (45): 28271-276. For example, to assay for a *Neisseria lgtC* enzyme, either FCHASE-AP-Lac or FCHASE-AP-Gal can be used, whereas for the *Neisseria lgtB* enzyme an appropriate reagent is FCHASE-AP-GlcNAc (*Id.*). Other methods for detection of a fucosylated reaction product include thin layer chromatography and GC/MS.

[0085] Also, a nucleic acid encoding an *H. pylori* α -1,3/4-fucosyltransferase, or a subsequence thereof, can be chemically synthesized. Suitable methods include the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68: 109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill recognizes that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

[0086] Nucleic acids encoding *H. pylori* α -1,3/4-fucosyltransferases, or subsequences thereof, can be cloned using DNA amplification methods such as polymerase chain reaction (PCR). Thus, for example, the nucleic acid sequence or subsequence is PCR amplified, using a sense primer containing one restriction enzyme site (e.g., *NdeI*) and an antisense primer containing another restriction enzyme site (e.g., *HindIII*). This will produce a nucleic acid encoding the desired *H. pylori* α -1,3/4-fucosyltransferases or subsequence and having terminal restriction enzyme sites. This nucleic acid can then be easily ligated into a vector containing a nucleic acid encoding the second molecule and having the appropriate corresponding restriction enzyme sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided in GenBank or other sources. Appropriate restriction enzyme sites can also be added to the nucleic acid encoding the *H. pylori* α -1,3/4-fucosyltransferase protein or protein subsequence by site-directed mutagenesis. The plasmid containing the *H. pylori* α -1,3/4-fucosyltransferase-encoding nucleotide sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.*, eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem.*, 35: 1826; Landegren *et al.*, (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117.

[0087] Other physical properties of a cloned *H. pylori* α -1,3/4-fucosyltransferase protein expressed from a particular nucleic acid, can be compared to properties of known *H. pylori* α -1,3/4-fucosyltransferases to provide another method of identifying suitable sequences or domains of the *H. pylori* α -1,3/4-fucosyltransferases that are determinants of acceptor substrate specificity and/or catalytic activity. Alternatively, a putative *H. pylori* α -1,3/4-fucosyltransferase gene or recombinant *H. pylori* α -1,3/4-fucosyltransferase gene can be mutated, and its role as an α -1,3/4-fucosyltransferases, or the role of particular sequences or domains established by detecting a variation in the structure of a carbohydrate normally produced by the unmutated, naturally-occurring, or control α -1,3/4-fucosyltransferases.

[0088] Functional domains of cloned *H. pylori* α -1,3/4-fucosyltransferases can be identified by using standard methods for mutating or modifying the g *H. pylori* α -1,3/4-fucosyltransferases and testing the modified or mutated proteins for activities such as acceptor substrate activity and/or catalytic activity, as described herein. The functional domains of the various *H. pylori* α -1,3/4-fucosyltransferases can be used to construct nucleic acids encoding α -1,3/4-fucosyltransferases proteins comprising the functional domains of one or more α -1,3/4-fucosyltransferases. These fusion proteins can then be tested for the desired acceptor substrate or catalytic activity.

[0089] In an exemplary approach to cloning nucleic acids encoding α -1,3/4-fucosyltransferase proteins, the known nucleic acid or amino acid sequences of cloned glycosyltransferases are aligned and compared to determine the amount of sequence identity between various glycosyltransferases. This information can be used to identify and select protein domains that confer or modulate glycosyltransferase activities, e.g., acceptor substrate activity and/or catalytic activity based on the amount of sequence identity between the glycosyltransferases of interest. For example, domains having sequence identity between the fucosyltransferases of interest, and that are associated with a known activity, can be used to construct fucosyltransferase proteins containing that domain, and having the activity associated with that domain (e.g., acceptor substrate specificity and/or catalytic activity).

B. Fusion protein comprising accessory enzymes involved in nucleotide sugar formation

[0090] In some embodiments, the fusion polypeptides of the invention include, in addition to the α -1,3/4-fucosyltransferases catalytic domain(s) and/or other functional domains, at least one catalytic domain from an accessory enzyme. Accessory enzymes include, for example, those enzymes that are involved in the formation of a nucleotide sugar. The accessory enzyme can be involved in attaching the sugar to a nucleotide, or can be involved in making the sugar or the nucleotide, for example. The nucleotide sugar is generally one that is utilized as a saccharide donor by the glycosyltransferase catalytic domain of the particular fusion polypeptide. α -1,3/4-fucosyltransferases utilize GDP-fucose as a sugar donor. Examples of fusion proteins comprising a functional domain from a glycosyltransferase and an accessory enzyme and methods to make such fusions are found for example in PCT/CA98/01180, USSN 09/211,691 filed December 14, 1998 both of which are herein incorporated by reference for all purposes.

[0091] Accessory enzymes that are involved in synthesis of nucleotide sugars are well known to those of skill in the art. For a review of bacterial polysaccharide synthesis and gene nomenclature, *see, e.g.,* Reeves *et al.*, *Trends Microbiol.* 4: 495-503 (1996). The methods described above for obtaining glycosyltransferase-encoding nucleic acids are also applicable to obtaining nucleic acids that encode enzymes involved in the formation of nucleotide sugars. For example, one can use one of nucleic acids known in the art, some of which are listed below, directly or as a probe to isolate a corresponding nucleic acid from other organisms of interest.

[0092] An example of a fusion polypeptide provided by the invention is used for producing a fucosylated soluble oligosaccharide. The donor nucleotide sugar for fucosyltransferases is GDP-fucose, which is relatively expensive to produce. To reduce the cost of producing the fucosylated oligosaccharide, the invention provides fusion polypeptides that can convert the relatively inexpensive GDP-mannose into GDP-fucose, and then catalyze the transfer of the fucose to an acceptor saccharide. These fusion polypeptides include a catalytic domain from at least one of a GDP-mannose dehydratase, a GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase, or a GDP-4-keto-6-deoxy-L-glucose 4-reductase. When each of these enzyme activities is provided, one can convert GDP-mannose into GDP-fucose.

[0093] The nucleotide sequence of an *E. coli* gene cluster that encodes GDP-fucose-synthesizing enzymes is described by Stevenson *et al.* (1996) *J. Bacteriol.* 178: 4885-4893; GenBank Accession No. U38473). This gene cluster had been reported to include an open reading frame for GDP-mannose dehydratase (nucleotides 8633-9754; Stevenson *et al.*, *supra.*). It was recently discovered that this gene cluster also contains an open reading frame that encodes an enzyme that has both 3,5 epimerization and 4-reductase activities (*see*, commonly assigned US Patent No. 6,500,661, issued December 31, 2002), and thus is capable of converting the product of the GDP-mannose dehydratase reaction (GDP-4-keto-6-deoxymannose) to GDP-fucose. This ORF, which is designated YEF B, is found between nucleotides 9757-10722. Prior to this discovery that YEF B encodes an enzyme having two activities, it was not known whether one or two enzymes were required for conversion of GDP-4-keto-6-deoxymannose to GDP-fucose. The nucleotide sequence of a gene encoding the human Fx enzyme is found in GenBank Accession No. U58766.

[0094] Also provided are fusion polypeptides that include a mannosyltransferase catalytic domain and a catalytic domain of a GDP-Man pyrophosphorylase (EC 2.7.7.22), which

converts Man-1-P to GDP-Man. Suitable genes are known from many organisms, including *E. coli*: GenBank U13629, AB010294, D43637 D13231, Bastin *et al.*, *Gene* 164: 17-23 (1995), Sugiyama *et al.*, *J. Bacteriol.* 180: 2775-2778 (1998), Sugiyama *et al.*, *Microbiology* 140 (Pt 1): 59-71 (1994), Kido *et al.*, *J. Bacteriol.* 177: 2178-2187 (1995); *Klebsiella pneumoniae*: GenBank AB010296, AB010295, Sugiyama *et al.*, *J. Bacteriol.* 180: 2775-2778 (1998); *Salmonella enterica*: GenBank X56793 M29713, Stevenson *et al.*, *J. Bacteriol.* 178: 4885-4893 (1996).

[0095] The fusion polypeptides of the invention for fucosylating a saccharide acceptor can also utilize enzymes that provide a minor or "scavenge" pathway for GDP-fucose formation.

In this pathway, free fucose is phosphorylated by fucokinase to form fucose 1-phosphate, which, along with guanosine 5'-triphosphate (GTP), is used by GDP-fucose pyrophosphorylase to form GDP-fucose (Ginsburg *et al.*, *J. Biol. Chem.*, 236: 2389-2393 (1961) and Reitman, *J. Biol. Chem.*, 255: 9900-9906 (1980)). Accordingly, a fucosyltransferase catalytic domain can be linked to a catalytic domain from a GDP-fucose pyrophosphorylase, for which suitable nucleic acids are described in copending, commonly assigned US Patent Application Ser. No. 08/826,964, filed April 9, 1997. Fucokinase-encoding nucleic acids are described for, *e.g.*, *Haemophilus influenzae* (Fleischmann *et al.* (1995) *Science* 269:496-512) and *E. coli* (Lu and Lin (1989) *Nucleic Acids Res.* 17: 4883-4884).

[0096] Additional accessory enzymes from which one can obtain a catalytic domain are those that are involved in forming reactants consumed in a glycosyltransferase cycle. For example, any of several phosphate kinases are useful as accessory enzymes. Polyphosphate kinase (EC 2.7.4.1), for example, catalyzes the formation of ATP; nucleoside phosphate kinases (EC 2.7.4.4) can form the respective nucleoside diphosphates; creatine phosphate kinase (EC 2.7.3.2); myokinase (EC 2.7.4.3); *N*-acetylglucosamine acetyl kinase (EC 2.7.1.59); acetyl phosphate kinase; and pyruvate kinase (EC 2.7.1.40).

C. Expression cassettes and host cells for expressing recombinant *H. pylori* α -1,3/4-fucosyltransferase proteins

[0097] Fusion proteins of the invention can be expressed in a variety of host cells, including *E. coli*, other bacterial hosts, and yeast. The host cells are preferably microorganisms, such as, for example, yeast cells, bacterial cells, or filamentous fungal cells. Examples of suitable host cells include, for example, *Azotobacter sp.* (*e.g.*, *A. vinelandii*),

Pseudomonas sp., *Rhizobium* sp., *Erwinia* sp., *Escherichia* sp. (e.g., *E. coli*), *Bacillus*, *Pseudomonas*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, *Paracoccus* and *Klebsiella* sp., among many others. The cells can be of any of several genera, including *Saccharomyces* (e.g., *S. cerevisiae*), *Candida* (e.g., *C. utilis*, *C. parapsilosis*, *C. krusei*, *C. versatilis*, *C. lipolytica*, *C. zeylanoides*, *C. guilliermondii*, *C. albicans*, and *C. humicola*), *Pichia* (e.g., *P. farinosa* and *P. ohmeri*), *Torulopsis* (e.g., *T. candida*, *T. sphaerica*, *T. xylinus*, *T. famata*, and *T. versatilis*), *Debaryomyces* (e.g., *D. subglobosus*, *D. cantarellii*, *D. globosus*, *D. hansenii*, and *D. japonicus*), *Zygosaccharomyces* (e.g., *Z. rouxii* and *Z. bailii*), *Kluyveromyces* (e.g., *K. marxianus*), *Hansenula* (e.g., *H. anomala* and *H. jadinii*), and *Brettanomyces* (e.g., *B. lambicus* and *B. anomalus*). Examples of useful bacteria include, but are not limited to, *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Klebsiella*.

[0098] Typically, the polynucleotide that encodes the α -1,3/4-fucosyltransferase protein is placed under the control of a promoter that is functional in the desired host cell. An extremely wide variety of promoters are well known, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include one or more of these control sequences are termed "expression cassettes." Accordingly, the invention provides expression cassettes into which the nucleic acids that encode fusion proteins are incorporated for high level expression in a desired host cell.

[0099] Expression control sequences that are suitable for use in a particular host cell are often obtained by cloning a gene that is expressed in that cell. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (*lac*) promoter systems (Change *et al.*, *Nature* (1977) 198: 1056), the tryptophan (*trp*) promoter system (Goeddel *et al.*, *Nucleic Acids Res.* (1980) 8: 4057), the *tac* promoter (DeBoer, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); and the lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake *et al.*, *Nature* (1981) 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes can be used.

[0100] For expression of α -1,3/4-fucosyltransferase proteins in prokaryotic cells other than *E. coli*, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, the hybrid *trp-lac* promoter functions in *Bacillus* in addition to *E. coli*.

[0101] A ribosome binding site (RBS) is conveniently included in the expression cassettes of the invention. An RBS in *E. coli*, for example, consists of a nucleotide sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine and Dalgarno, *Nature* (1975) 254: 34; Steitz, *In Biological regulation and development: Gene expression* (ed. R.F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

[0102] For expression of the α -1,3/4-fucosyltransferase proteins in yeast, convenient promoters include GAL1-10 (Johnson and Davies (1984) *Mol. Cell. Biol.* 4:1440-1448) ADH2 (Russell *et al.* (1983) *J. Biol. Chem.* 258:2674-2682), PHO5 (*EMBO J.* (1982) 6:675-680), and MF α (Herskowitz and Oshima (1982) in *The Molecular Biology of the Yeast Saccharomyces* (eds. Strathern, Jones, and Broach) Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., pp. 181-209). Another suitable promoter for use in yeast is the ADH2/GAPDH hybrid promoter as described in Cousens *et al.*, *Gene* 61:265-275 (1987). For filamentous fungi such as, for example, strains of the fungi *Aspergillus* (McKnight *et al.*, U.S. Patent No. 4,935,349), examples of useful promoters include those derived from *Aspergillus nidulans* glycolytic genes, such as the ADH3 promoter (McKnight *et al.*, *EMBO J.* 4: 2093 2099 (1985)) and the *tpiA* promoter. An example of a suitable terminator is the ADH3 terminator (McKnight *et al.*).

[0103] Either constitutive or regulated promoters can be used in the present invention. Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the fusion proteins is induced. High level expression of heterologous proteins slows cell growth in some situations. An inducible promoter is a promoter that directs expression of a gene where the level of expression is alterable by environmental or developmental factors such as, for example, temperature, pH, anaerobic or aerobic conditions, light, transcription factors and chemicals. Such promoters are referred to herein as "inducible" promoters, which allow one to control the timing of expression of the glycosyltransferase or enzyme involved in nucleotide sugar synthesis. For *E. coli* and other bacterial host cells, inducible promoters are known to those of skill in the art. These include,

for example, the *lac* promoter, the bacteriophage lambda P_L promoter, the hybrid *trp-lac* promoter (Amann *et al.* (1983) *Gene* 25: 167; de Boer *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA* 80: 21), and the bacteriophage T7 promoter (Studier *et al.* (1986) *J. Mol. Biol.*; Tabor *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA* 82: 1074-8). These promoters and their use are

discussed in Sambrook *et al.*, *supra*. A particularly preferred inducible promoter for expression in prokaryotes is a dual promoter that includes a *tac* promoter component linked to a promoter component obtained from a gene or genes that encode enzymes involved in galactose metabolism (*e.g.*, a promoter from a UDPgalactose 4-epimerase gene (*galE*)). The dual *tac-gal* promoter, which is described in PCT Patent Application Publ. No. WO98/20111,

[0104] A construct that includes a polynucleotide of interest operably linked to gene expression control signals that, when placed in an appropriate host cell, drive expression of the polynucleotide is termed an "expression cassette." Expression cassettes that encode the fusion proteins of the invention are often placed in expression vectors for introduction into the host cell. The vectors typically include, in addition to an expression cassette, a nucleic acid sequence that enables the vector to replicate independently in one or more selected host cells. Generally, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria. For instance, the origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria.

Alternatively, the vector can replicate by becoming integrated into the host cell genomic complement and being replicated as the cell undergoes DNA replication. A preferred expression vector for expression of the enzymes is in bacterial cells is pTGK, which includes a dual *tac-gal* promoter and is described in PCT Patent Application Publ. NO. WO98/20111.

[0105] The construction of polynucleotide constructs generally requires the use of vectors able to replicate in bacteria. A plethora of kits are commercially available for the purification of plasmids from bacteria (*see*, for example, EasyPrepJ, FlexiPrepJ, both from Pharmacia Biotech; StrataCleanJ, from Stratagene; and, QIAexpress Expression System, Qiagen). The isolated and purified plasmids can then be further manipulated to produce other plasmids, and used to transfect cells. Cloning in *Streptomyces* or *Bacillus* is also possible.

[0106] Selectable markers are often incorporated into the expression vectors used to express the polynucleotides of the invention. These genes can encode a gene product, such as a protein, necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will

not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, such as ampicillin, neomycin, kanamycin, chloramphenicol, or tetracycline. Alternatively, selectable markers may encode proteins that complement auxotrophic deficiencies or supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for Bacilli. Often, the vector will have one selectable marker that is functional in, *e.g.*, *E. coli*, or other cells in which the vector is replicated prior to being introduced into the host cell. A number of selectable markers are known to those of skill in the art and are described for instance in Sambrook *et al.*, *supra*.

[0107] Construction of suitable vectors containing one or more of the above listed components employs standard ligation techniques as described in the references cited above. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required. To confirm correct sequences in plasmids constructed, the plasmids can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequencing according to known methods. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Volume 152, Academic Press, Inc., San Diego, CA (Berger); and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement) (Ausubel).

[0108] A variety of common vectors suitable for use as starting materials for constructing the expression vectors of the invention are well known in the art. For cloning in bacteria, common vectors include pBR322 derived vectors such as pBLUESCRIPT™, and λ -phage derived vectors. In yeast, vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp series plasmids) and pGPD-2. Expression in mammalian cells can be achieved using a variety of commonly available plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses).

[0109] The methods for introducing the expression vectors into a chosen host cell are not particularly critical, and such methods are known to those of skill in the art. For example, the expression vectors can be introduced into prokaryotic cells, including *E. coli*, by calcium chloride transformation, and into eukaryotic cells by calcium phosphate treatment or electroporation. Other transformation methods are also suitable.

[0110] Translational coupling may be used to enhance expression. The strategy uses a short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. See Squires, et. al. (1988), *J. Biol. Chem.* 263: 16297-16302.

[0111] The α -1,3/4-fucosyltransferase proteins can be expressed intracellularly, or can be secreted from the cell. Intracellular expression often results in high yields. If necessary, the amount of soluble, active fusion protein may be increased by performing refolding procedures (see, e.g., Sambrook et al., *supra.*; Marston et al., *Bio/Technology* (1984) 2: 800; Schoner et al., *Bio/Technology* (1985) 3: 151). In embodiments in which the α -1,3/4-fucosyltransferase proteins are secreted from the cell, either into the periplasm or into the extracellular medium, the DNA sequence is linked to a cleavable signal peptide sequence. The signal sequence directs translocation of the fusion protein through the cell membrane. An example of a suitable vector for use in *E. coli* that contains a promoter-signal sequence unit is pTA1529, which has the *E. coli* *phoA* promoter and signal sequence (see, e.g., Sambrook et al., *supra.*; Oka et al., *Proc. Natl. Acad. Sci. USA* (1985) 82: 7212; Talmadge et al., *Proc. Natl. Acad. Sci. USA* (1980) 77: 3988; Takahara et al., *J. Biol. Chem.* (1985) 260: 2670). In another embodiment, the fusion proteins are fused to a subsequence of protein A or bovine serum albumin (BSA), for example, to facilitate purification, secretion, or stability.

[0112] The α -1,3/4-fucosyltransferase proteins of the invention can also be further linked to other bacterial proteins. This approach often results in high yields, because normal prokaryotic control sequences direct transcription and translation. In *E. coli*, *lacZ* fusions are often used to express heterologous proteins. Suitable vectors are readily available, such as the pUR, pEX, and pMR100 series (see, e.g., Sambrook et al., *supra.*). For certain

applications, it may be desirable to cleave the non-glycosyltransferase and/or accessory enzyme amino acids from the fusion protein after purification. This can be accomplished by any of several methods known in the art, including cleavage by cyanogen bromide, a protease, or by Factor X_a (*see, e.g.,* Sambrook *et al., supra.*; Itakura *et al., Science* (1977) 198: 1056; Goeddel *et al., Proc. Natl. Acad. Sci. USA* (1979) 76: 106; Nagai *et al., Nature* (1984) 309: 810; Sung *et al., Proc. Natl. Acad. Sci. USA* (1986) 83: 561). Cleavage sites can be engineered into the gene for the fusion protein at the desired point of cleavage.

[0113] More than one recombinant protein may be expressed in a single host cell by placing multiple transcriptional cassettes in a single expression vector, or by utilizing different selectable markers for each of the expression vectors which are employed in the cloning strategy.

[0114] A suitable system for obtaining recombinant proteins from *E. coli* which maintains the integrity of their N-termini has been described by Miller *et al. Biotechnology* 7:698-704 (1989). In this system, the gene of interest is produced as a C-terminal fusion to the first 76 residues of the yeast ubiquitin gene containing a peptidase cleavage site. Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal residue.

D. Purification of α -1,3/4-fucosyltransferase proteins

[0115] The *H. pylori* fucosyltransferase proteins of the present invention can be expressed as intracellular proteins or as proteins that are secreted from the cell, and can be used in this form, in the methods of the present invention. For example, a crude cellular extract containing the expressed intracellular or secreted *H. pylori* fucosyltransferase protein can be used in the methods of the present invention.

[0116] Alternatively, the *H. pylori* fucosyltransferase proteins can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see, generally, R. Scopes, Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). Substantially pure compositions of at least about 70 to 90% homogeneity are preferred, and 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.,* as immunogens for antibody production.

[0117] To facilitate purification of the *H. pylori* α -1,3/4-fucosyltransferase proteins of the invention, the nucleic acids that encode the fusion proteins can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available, *i.e.* a purification tag. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion proteins having these epitopes are commercially available (*e.g.*, Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the *H. pylori* α -1,3/4-fucosyltransferase proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (*e.g.*, FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines are used, although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding moiety for a polyhistidine tag include nitrilo-tri-acetic acid (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" In Genetic Engineering: Principles and Methods, J.K. Setlow, Ed., Plenum Press, NY; commercially available from Qiagen (Santa Clarita, CA)).

[0118] Purification tags also include maltose binding domains and starch binding domains. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacylodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0119] Other haptens that are suitable for use as tags are known to those of skill in the art and are described, for example, in the Handbook of Fluorescent Probes and Research Chemicals (6th Ed., Molecular Probes, Inc., Eugene OR). For example, dinitrophenol (DNP), digoxigenin, barbiturates (see, *e.g.*, US Patent No. 5,414,085), and several types of fluorophores are useful as haptens, as are derivatives of these compounds. Kits are commercially available for linking haptens and other moieties to proteins and other molecules. For example, where the hapten includes a thiol, a heterobifunctional linker such as SMCC can be used to attach the tag to lysine residues present on the capture reagent.

[0120] One of skill would recognize that modifications can be made to the α -1,3/4-fucosyltransferase catalytic or functional domains without diminishing their biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the catalytic domain into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, the addition of codons at either terminus of the polynucleotide that encodes the catalytic domain to provide, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located restriction enzyme sites or termination codons or purification sequences.

E. Uses of the *H. pylori* fucosyltransferase proteins

[0121] The invention provides *H. pylori* α -1,3/4-fucosyltransferase proteins and methods of using the *H. pylori* α -1,3/4-fucosyltransferase proteins to enzymatically synthesize glycolipids, and oligosaccharide moieties. The glycosyltransferase reactions of the invention take place in a reaction medium comprising at least one *H. pylori* α -1,3/4-fucosyltransferase, acceptor substrate, and donor substrate, and typically a soluble divalent metal cation. In some embodiments, accessory enzymes and substrates for the accessory enzyme catalytic moiety are also present, so that the accessory enzymes can synthesize the donor substrate for the *H. pylori* α -1,3/4-fucosyltransferase.

[0122] The *H. pylori* fucosyltransferase proteins prepared as described herein can be used in combination with additional glycosyltransferases. For example, one can use a combination of recombinant sialyltransferase fusion protein and a recombinant *H. pylori* α -1,3/4-fucosyltransferases. By conducting two glycosyltransferase reactions in sequence in a single vessel, overall yields are improved over procedures in which an intermediate species is isolated. Moreover, cleanup and disposal of extra solvents and by-products is reduced.

Similarly, the recombinant glycosyltransferases can be used with recombinant accessory enzyme, which may or may not be present as a the fusion protein. In other embodiments, the *H. pylori* α -1,3/4-fucosyltransferase and additional glycosyltransferases or accessory enzymes are produced in the same cell and used to synthesize a desired end product.

[0123] The products produced by the above processes can be used without purification.

However, standard, well known techniques, for example, thin or thick layer chromatography, ion exchange chromatography, or membrane filtration can be used for recovery of glycosylated saccharides. Also, for example, membrane filtration, utilizing a nanofiltration

or reverse osmotic membrane as described in commonly assigned AU Patent No. 735695 may be used. As a further example, membrane filtration wherein the membranes have a molecular weight cutoff of about 1000 to about 10,000 Daltons can be used to remove proteins. As another example, nanofiltration or reverse osmosis can then be used to remove salts. Nanofilter membranes are a class of reverse osmosis membranes which pass monovalent salts but retain polyvalent salts and uncharged solutes larger than about 200 to about 1000 Daltons, depending upon the membrane used. Thus, for example, the oligosaccharides produced by the compositions and methods of the present invention can be retained in the membrane and contaminating salts will pass through.

F. Donor Substrates and Acceptor Substrates

[0124] Suitable donor substrates used by the *H. pylori* fucosyltransferase proteins and other glycosyltransferases in the methods of the invention include, but are not limited to, UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, GDP-Man, GDP-Fuc, UDP-GlcUA, and CMP-sialic acid. Guo *et al.*, *Applied Biochem. and Biotech.* **68**: 1-20 (1997)

[0125] Suitable acceptor substrates used by the *H. pylori* fucosyltransferase proteins and methods of the invention include, but are not limited to, polysaccharides, oligosaccharides, lipids, and glycolipids. For example, the oligosaccharide LNnT can be fucosylated to form LNFI. The fucosyltransferases described herein can also be used in multienzyme systems to produce a desired product from a convenient starting material. For example, LNFI was prepared on a multigram scale from lactose using the *H. pylori* α -1,3/4-fucosyltransferases from strain 1182 described herein, in combination with *Neisseria gonococcus* β -1,3N-acetylglucosaminyltransferase (lgtA) and *Neisseria gonococcus* β -1,4-galactosyltransferase (lgtB).

[0126] The recombinant fusion protein of the invention can be constructed and expressed as a fusion protein with a molecular "tag" at one end, which facilitates purification of the protein, *i.e.*, a purification tag. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include "epitope tags," which are a protein sequence that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A "FLAG tag" is a commonly used epitope tag, specifically recognized by a monoclonal anti-FLAG antibody, consisting of the sequence AspTyrLysAspAspAsp AspLys or a substantially identical variant thereof. A myc

tag is another commonly used epitope tag. Other suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide, which will bind to metal ions such as nickel or cobalt ions. Purification tags also include maltose binding domains and starch binding domains. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacylodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

G. Fucosyltransferase reactions

[0127] The *H. pylori* fucosyltransferase proteins, acceptor substrates, donor substrates and other reaction mixture ingredients, including other glycosyltransferases and accessory enzymes are combined by admixture in an aqueous reaction medium. The medium generally has a pH value of about 5 to about 8.5. The selection of a medium is based on the ability of the medium to maintain pH value at the desired level. Thus, in some embodiments, the medium is buffered to a pH value of about 7.5. If a buffer is not used, the pH of the medium should be maintained at about 5 to 8.5, depending upon the particular glycosyltransferase used. For fucosyltransferases, the pH range is preferably maintained from about 6.0 to 8.0. For sialyltransferases, the range is preferably from about 5.5 and about 7.5.

[0128] Enzyme amounts or concentrations are expressed in activity units, which is a measure of the initial rate of catalysis. One activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (typically 37°C) and pH value (typically 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of 37 °C and a pH value of 7.5.

[0129] The reaction mixture may include divalent metal cations (Mg^{2+} , Mn^{2+}). The reaction medium may also comprise solubilizing detergents (e.g., Triton or SDS) and organic solvents such as methanol or ethanol, if necessary. The enzymes can be utilized free in solution or can be bound to a support such as a polymer. The reaction mixture is thus substantially homogeneous at the beginning, although some precipitate can form during the reaction.

[0130] The temperature at which an above process is carried out can range from just above freezing to the temperature at which the most sensitive enzyme denatures. That temperature range is preferably about 0°C to about 45°C, and more preferably at about 20°C to about 37°C.

5 [0131] The reaction mixture so formed is maintained for a period of time sufficient to obtain the desired high yield of desired oligosaccharide products, including determinants present on oligosaccharide groups attached to the glycolipid to be glycosylated. For large-scale preparations, the reaction will often be allowed to proceed for between about 0.5-240 hours, and more typically between about 1-18 hours.

10 [0132] In embodiments in which more than one glycosyltransferase is used to obtain the oligosaccharide products, the enzymes and reagents for a second glycosyltransferase reaction can be added to the reaction medium once the first glycosyltransferase reaction has neared completion. For some combinations of enzymes, the glycosyltransferases and corresponding substrates can be combined in a single initial reaction mixture; the enzymes in such
15 simultaneous reactions preferably do not form a product that cannot serve as an acceptor for the other enzyme. By conducting two glycosyltransferase reactions in sequence in a single vessel, overall yields are improved over procedures in which an intermediate species is isolated. Moreover, cleanup and disposal of extra solvents and by-products is reduced. In addition, in some embodiments, the fucosyltransferase and additionally glycosyltransferases
20 or accessory enzymes are expressed in the same host cell and the desired product is synthesized within the host cell.

[0133] One or more of the glycosyltransferase reactions can be carried out as part of a glycosyltransferase cycle. Preferred conditions and descriptions of glycosyltransferase cycles have been described. A number of glycosyltransferase cycles (for example, sialyltransferase
25 cycles, galactosyltransferase cycles, and fucosyltransferase cycles) are described in U.S. Patent No. 5,374,541 and WO 9425615 A. Other glycosyltransferase cycles are described in Ichikawa *et al.* *J. Am. Chem. Soc.* 114:9283 (1992), Wong *et al.* *J. Org. Chem.* 57: 4343 (1992), DeLuca, *et al.*, *J. Am. Chem. Soc.* 117:5869-5870 (1995), and Ichikawa *et al.* In *Carbohydrates and Carbohydrate Polymers*. Yaltami, ed. (ATL Press, 1993).

30 [0134] Other glycosyltransferases can be substituted into similar transferase cycles as have been described in detail for the fucosyltransferases and sialyltransferases. In particular, the glycosyltransferase can also be, for instance, glucosyltransferases, *e.g.*, Alg8 (Stagljev *et al.*,

Proc. Natl. Acad. Sci. USA 91:5977 (1994)) or Alg5 (Heesen *et al. Eur. J. Biochem.* 224:71 (1994)), N-acetylgalactosaminyltransferases such as, for example, $\alpha(1,3)$ N-acetylgalactosaminyltransferase, $\beta(1,4)$ N-acetylgalactosaminyltransferases (Nagata *et al. J. Biol. Chem.* 267:12082-12089 (1992) and Smith *et al. J. Biol. Chem.* 269:15162 (1994)) and polypeptide N-acetylgalactosaminyltransferase (Homa *et al. J. Biol. Chem.* 268:12609 (1993)). Suitable N-acetylglucosaminyltransferases include GnTI (2.4.1.101, Hull *et al., BBRC* 176:608 (1991)), GnTII, and GnTIII (Ihara *et al. J. Biochem.* 113:692 (1993)), GnTV (Shoreiban *et al. J. Biol. Chem.* 268: 15381 (1993)), O-linked N-acetylglucosaminyltransferase (Bierhuizen *et al. Proc. Natl. Acad. Sci. USA* 89:9326 (1992)), N-acetylglucosamine-1-phosphate transferase (Rajput *et al. Biochem J.* 285:985 (1992), and hyaluronan synthase. Suitable mannosyltransferases include $\alpha(1,2)$ mannosyltransferase, $\alpha(1,3)$ mannosyltransferase, $\beta(1,4)$ mannosyltransferase, Dol-P-Man synthase, OCh1, and Pmt1.

[0135] For the above glycosyltransferase cycles, the concentrations or amounts of the various reactants used in the processes depend upon numerous factors including reaction conditions such as temperature and pH value, and the choice and amount of acceptor saccharides to be glycosylated. Because the glycosylation process permits regeneration of activating nucleotides, activated donor sugars and scavenging of produced PPi in the presence of catalytic amounts of the enzymes, the process is limited by the concentrations or amounts of the stoichiometric substrates discussed before. The upper limit for the concentrations of reactants that can be used in accordance with the method of the present invention is determined by the solubility of such reactants.

[0136] Preferably, the concentrations of activating nucleotides, phosphate donor, the donor sugar and enzymes are selected such that glycosylation proceeds until the acceptor is consumed.

[0137] Each of the enzymes is present in a catalytic amount. The catalytic amount of a particular enzyme varies according to the concentration of that enzyme's substrate as well as to reaction conditions such as temperature, time and pH value. Means for determining the catalytic amount for a given enzyme under preselected substrate concentrations and reaction conditions are well known to those of skill in the art.

[0138] The fucosyltransferase reaction can be carried out using an oligosaccharide or polysaccharide as an acceptor molecule. Suitable acceptor substrates used by the *H. pylori*

fucosyltransferase proteins and methods of the invention include, but are not limited to, polysaccharides, oligosaccharides, lipids, and glycolipids. For example, the oligosaccharide LNnT can be fucosylated to form LNFIH. The fucosyltransferases described herein can also be used in multienzyme systems to produce a desired product from a convenient starting material. For example, LNFIH was prepared on a multigram scale from lactose using the *H. pylori* α -1,3/4-fucosyltransferases from strain 1182 described herein, in combination with *Neisseria gonococcus* β -1,3N-acetylglucosaminyltransferase (lgtA) and *Neisseria gonococcus* β -1,4-galactosyltransferase (lgtB).

[0139] The recombinant fucosyltransferase fusion protein used in the methods of the invention is chosen based upon its ability to fucosylate the fucosyltransferase acceptor substrates of interest. Preferably, the fucosyltransferase is assayed for suitability using a fucosyltransferase acceptor substrate that is attached to a soluble saccharide or oligosaccharide. The use of a soluble saccharide or oligosaccharide acceptor substrate in the assay to determine fucosyltransferase activity allows one to select a fucosyltransferase that produces the desired oligosaccharide product.

[0140] The fucosyltransferase reaction can be carried out using a lipid or glycolipid as an acceptor molecule. Many saccharides require the presence of particular fucosylated structures in order to exhibit biological activity. Intercellular recognition mechanisms often require a fucosylated oligosaccharide. For example, a number of proteins that function as cell adhesion molecules, including P-selectin, E-selectin, bind specific cell surface fucosylated carbohydrate structures, for example, the sialyl Lewis x and the sialyl Lewis a structures. In addition, the specific carbohydrate structures that form the ABO blood group system are fucosylated. The carbohydrate structures in each of the three groups share a Fuc α 1,2Gal β 1-dissacharide unit. In blood group O structures, this disaccharide is the terminal structure. The group A structure is formed by an α 1,3 GalNAc transferase that adds a terminal GalNAc residue to the dissacharide. The group B structure is formed by an α 1,3 galactosyltransferase that adds terminal galactose residue. The Lewis blood group structures are also fucosylated. For example the Lewis x and Lewis a structures are Gal β 1,4(Fuc α 1,3)GlcNAc and Gal β 1,4(Fuc α 1,4)GlcNAc, respectively. Both these structures can be further sialylated (NeuAc α 2,3-) to form the corresponding sialylated structures. Other Lewis blood group structures of interest are the Lewis y and b structures which are Fuc α 1,2Gal β 1,4(Fuc α 1,3)GlcNAc β -OR and Fuc α 1,2Gal β 1,3(Fuc α 1,4)GlcNAc-OR,

respectively. For a description of the structures of the ABO and Lewis blood group structures and the enzymes involved in their synthesis *see, Essentials of Glycobiology*, Varki *et al.* eds., Chapter 16 (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1999).

[0141] The recombinant fucosyltransferase fusion protein used in the methods of the invention is chosen based upon its ability to fucosylate the fucosyltransferase acceptor substrates of interest. Preferably, the fucosyltransferase is assayed for suitability using a fucosyltransferase acceptor substrate that is attached to a lipid or glycolipid. The use of a glycolipid-linked acceptor substrate, rather than an acceptor substrate that is part of a soluble oligosaccharide, in the assay to determine fucosyltransferase activity allows one to select a fucosyltransferase that produces the selected fucosylation pattern on the glycolipid.

[0142] Fucosyltransferases have been used in synthetic pathways to transfer a fucose unit from guanosine-5'-diphosphofucose to a specific hydroxyl of a saccharide acceptor. For example, Ichikawa prepared sialyl Lewis-X by a method that involves the fucosylation of sialylated lactosamine with a cloned fucosyltransferase (Ichikawa *et al.*, *J. Am. Chem. Soc.* **114**: 9283-9298 (1992)). Lowe has described a method for expressing non-native fucosylation activity in cells, thereby producing fucosylated glycoproteins, cell surfaces, *etc.* (U.S. Patent No. 5,955,347).

[0143] In one embodiment, the methods of the invention are practiced by contacting a substrate, having an acceptor moiety for a fucosyltransferase, with a reaction mixture that includes a fucose donor moiety, a fucosyltransferase, and other reagents required for fucosyltransferase activity. The substrate is incubated in the reaction mixture for a sufficient time and under appropriate conditions to transfer fucose from the fucose donor moiety to the fucosyltransferase acceptor moiety. In preferred embodiments, the fucosyltransferase catalyzes the fucosylation of at least 60% of the fucosyltransferase respective acceptor moieties in the composition.

[0144] Specificity for a selected substrate is only the first criterion a preferred fucosyltransferase should satisfy. The fucosyltransferase used in the method of the invention is preferably also able to efficiently fucosylate a variety of substrates, and support scale-up of the reaction to allow the fucosylation of at least about 500 mg of the substrate. More preferably, the fucosyltransferase will support the scale of the fucosylation reaction to allow the synthesis of at least about 1 kg, and more preferably, at least 10 kg of substrate with relatively low cost and infrastructure requirements.

[0145] Suitable acceptor moieties for fucosyltransferase-catalyzed attachment of a fucose residue include, but are not limited to, GlcNAc-OR, Gal β 1,3GlcNAc-OR, NeuAc α 2,3Gal β 1,3GlcNAc-OR, Gal β 1,4GlcNAc-OR and NeuAc α 2,3Gal β 1,4GlcNAc-OR, where R is an amino acid, a saccharide, an oligosaccharide or an aglycon group having at least one carbon atom. R is linked to or is part of a substrate. The appropriate fucosyltransferase for a particular reaction is chosen based on the type of fucose linkage that is desired (*e.g.*, α 2, α 3, or α 4), the particular acceptor of interest, and the ability of the fucosyltransferase to achieve the desired high yield of fucosylation. Suitable fucosyltransferases and their properties are described above.

[0146] If a sufficient proportion of the substrate-linked oligosaccharides in a composition does not include a fucosyltransferase acceptor moiety, one can synthesize a suitable acceptor. For example, one preferred method for synthesizing an acceptor for a fucosyltransferase involves use of a GlcNAc transferase to attach a GlcNAc residue to a GlcNAc transferase acceptor moiety, which is present on the substrate-linked oligosaccharides. In preferred embodiments a transferase is chosen, having the ability to glycosylate a large fraction of the potential acceptor moieties of interest. The resulting GlcNAc β -OR can then be used as an acceptor for a fucosyltransferase.

[0147] The resulting GlcNAc β -OR moiety can be galactosylated prior to the fucosyltransferase reaction, yielding, for example, a Gal β 1,3GlcNAc-OR or Gal β 1,4GlcNAc-OR residue. In some embodiments, the galactylation and fucosylation steps can be carried out simultaneously. By choosing a fucosyltransferase that requires the galactosylated acceptor, only the desired product is formed. Thus, this method involves:

(a) galactosylating a compound of the formula GlcNAc β -OR with a galactosyltransferase in the presence of a UDP-galactose under conditions sufficient to form the compounds Gal β 1,4GlcNAc β -OR or Gal β 1,3GlcNAc-OR; and

(b) fucosylating the compound formed in (a) using a fucosyltransferase in the presence of GDP-fucose under conditions sufficient to form a compound selected from:

Fuc α 1,2Gal β 1,4GlcNAc1 β -O1R;

Fuc α 1,2Gal β 1,3GlcNAc-OR;

Fuc α 1,2Gal β 1,4GalNAc1 β -O1R;

Fuc α 1,2Gal β 1,3GalNAc-OR;

Gal β 1,4(Fuc1, α 3)GlcNAc β -OR; or
Gal β 1,3(Fuc α 1,4)GlcNAc-OR.

[0148] One can add additional fucose residues to the above structures by including an additional fucosyltransferase, which has the desired activity. For example, the methods can form oligosaccharide determinants such as Fuc α 1,2Gal β 1,4(Fuc α 1,3)GlcNAc β -OR and Fuc α 1,2Gal β 1,3(Fuc α 1,4)GlcNAc-OR. Thus, in another preferred embodiment, the method includes the use of at least two fucosyltransferases. The multiple fucosyltransferases are used either simultaneously or sequentially. When the fucosyltransferases are used sequentially, it is generally preferred that the glycoprotein is not purified between the multiple fucosylation steps. When the multiple fucosyltransferases are used simultaneously, the enzymatic activity can be derived from two separate enzymes or, alternatively, from a single enzyme having more than one fucosyltransferase activity.

H. Multiple-enzyme oligosaccharide synthesis

[0149] As discussed above, in some embodiments, two or more enzymes may be used to form a desired oligosaccharide or oligosaccharide determinant on a glycolipid. For example, a particular oligosaccharide determinant might require addition of a galactose, a sialic acid, and a fucose in order to exhibit a desired activity. Accordingly, the invention provides methods in which two or more enzymes, *e.g.*, glycosyltransferases, trans-sialidases, or sulfotransferases, are used to obtain high-yield synthesis of a desired oligosaccharide determinant.

[0150] In one preferred embodiment, LNFIII was prepared from lactose using the *H. pylori* α -1,3/4-fucosyltransferases from strain 1182 described herein, in combination with *Neisseria gonococcus* β -1,3N-acetylglucosaminyltransferase (lgtA) and *Neisseria gonococcus* β -1,4-galactosyltransferase (lgtB). Those of skill will recognize that other β -1,3N-acetylglucosaminyltransferase and β -1,4-galactosyltransferase enzymes can be used in this embodiment of the invention.

[0151] In some cases, a glycoprotein- or glycolipid linked oligosaccharide will include an acceptor substrate for the particular glycosyltransferase of interest upon *in vivo* biosynthesis of the glycoprotein or glycolipid. Such glycoproteins or glycolipids can be glycosylated using the *H. pylori* fucosyltransferase proteins and methods of the invention without prior modification of the glycosylation pattern of the glycoprotein or glycolipid, respectively. In other cases, however, a glycoprotein or glycolipid of interest will lack a suitable acceptor

substrate. In such cases, the methods of the invention can be used to alter the glycosylation pattern of the glycoprotein or glycolipid so that the glycoprotein-or glycolipid-linked oligosaccharides then include an acceptor substrate for the glycosyltransferase-catalyzed attachment of a preselected saccharide unit of interest to form a desired oligosaccharide moiety.

[0152] Glycolipid linked oligosaccharides optionally can be first “trimmed,” either in whole or in part, to expose either an acceptor substrate for the glycosyltransferase or a moiety to which one or more appropriate residues can be added to obtain a suitable acceptor substrate. Enzymes such as glycosyltransferases and endoglycosidases are useful for the attaching and trimming reactions. For example, a glycoprotein that displays “high mannose”-type oligosaccharides can be subjected to trimming by a mannosidase to obtain an acceptor substrate that, upon attachment of one or more preselected saccharide units, forms the desired oligosaccharide determinant.

[0153] The methods are also useful for synthesizing a desired oligosaccharide moiety on a protein or lipid that is unglycosylated in its native form. A suitable acceptor substrate for the corresponding glycosyltransferase can be attached to such proteins or lipids prior to glycosylation using the methods of the present invention. *See, e.g.*, US Patent No. 5,272,066 for methods of obtaining polypeptides having suitable acceptors for glycosylation.

[0154] Thus, in some embodiments, the invention provides methods for *in vitro* sialylation of saccharide groups present on a glycoconjugate that first involves modifying the glycoconjugate to create a suitable acceptor. Examples of preferred methods of multi-enzyme synthesis of desired oligosaccharide moieties are as follows.

EXAMPLES

[0155] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Cloning of *Helicobacter pylori* fucosyltransferases.

[0156] Putative fucosyltransferase genes from the following strains of *Helicobacter pylori* were PCR amplified, cloned into vectors for expression in *E. coli*: strain 915 FutA, strain 1111 FutA, strain 19C2 FutB, strain 1182 FutB, strain 19C2 FutA, strain 26695 FutA, and strain 1218 FutB. Nucleic acid and amino acid sequences are provided in Figures 1-7. An amino acid sequence alignment is provided in Figure 12; a nucleic acid sequence alignment is provided in Figure 13.

[0157] The putative fucosyltransferase proteins were screened for α 1,3/4-fucosyltransferase activity using LNnT and GDP-fucose substrates. The oligostructures of LNnT and one product, LNFPIII are shown in Figure 14.

[0158] One hundred milliliter cultures of *E. coli* transformed with *H. pylori*

5 fucosyltransferase were grown to OD600 of 0.8 and induced with IPTG, and harvested. Cell lysates were made using a french press. The fucosyltransferase enzymes were tested for enzymatic activity and acceptor specificity using the substrate LNnT. The reactions contained 3mM GDP-fucose, 3mM LNnT, 50mM Tris pH 7.5, 20mM MnCl₂, and 15% bacterial lysate. Reactions were incubated at 37°C for twenty-four hours.

10 [0159] Reaction products were separated using the following TLC- buffer system: 7 IPA:2 H₂O:1 Acetic acid. The samples were methylated, hydrolyzed, reduced with sodium borodeuteride, acetylated and analyzed by GC/MS along with samples of LNnT and LNF3. Results are shown in Figure 15. A Glc vs. Glc-NAc value close to 1 favors fucosylation of Glc-NAc. A Glc vs. Glc-NAc value close to 0 favors fucosylation of Glc.

15 Fucosyltransferases from the following *H. pylori* strains transferred fucose to Glc-NAc: strain 915 FutA, strain 1111 FutA, strain 19C2 FutB, and strain 1182 FutB. The FutA gene product from *H. pylori* strain 19C2A transferred fucose to the reducing glucose of the LNnT acceptor, as did the FutB gene product from *H. pylori* strain 1218 FutB. A novel FutA gene product from *H. pylori* strain 26695 also catalyzed the transfer of fucose to glucose.

20 Example 2: Production of oligosaccharides using *Helicobacter pylori* fucosyltransferases.

[0160] One liter cultures of *E. coli* expressing *H. pylori* fucosyltransferases were grown, induce and harvested. The lysates were used to synthesize LNFIII from LNnT. Two different ion exchange resins were tested for purification of LNFIII. Reaction mixtures were centrifuged at 5,000 RPM for thirty minutes. Samples were then processed by ultrafiltration
25 using hollow fiber ultrafiltration membranes with a molecular weight cut off of 10 kD. Ion exchange chromatography was done using either MR3 NH₄HCO₃ column 1ml resin per 1ml synthesis (70%) or Dowex1/Dowex50 column 2ml resin per 1ml synthesis (82%). Samples were then run on a P2 Size Exclusion column and then lyophilized. Results are shown in Figure 16. Yields using the Dowex resin approached 50%, while yields from the MR3
30 NH₄HCO₃ column approached 70%.

[0161] LNFIII was prepared from lactose using lysates from *E. coli* cells expressing *H. pylori* α -1,3/4-fucosyltransferases from strain 1182 described herein, in combination with

Neisseria gonococcus β -1,3N-acetylglucosaminyltransferase (lgtA) and *Neisseria gonococcus* β -1,4-galactosyltransferase (lgtB) on a multigram scale. Those of skill will recognize that other β -1,3N-acetylglucosaminyltransferase and β -1,4-galactosyltransferase enzymes can be used in this embodiment of the invention.

- 5 **[0162]** It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all
- 10 purposes

WHAT IS CLAIMED IS:

- 1 1. An isolated polynucleotide comprising a nucleic acid sequence,
2 wherein the nucleic acid sequence has greater than 90% identity to a nucleotide sequence
3 selected from the group consisting of SEQ ID NO:1, 3, 5, and 7, wherein the nucleotide
4 sequence encodes a fucosyltransferase that catalyzes the transfer of a fucose residue from a
5 donor substrate to an acceptor substrate.
- 1 2. The polynucleotide of claim 1, wherein the nucleic acid sequence is
2 selected from the group consisting of SEQ ID NO:1, 3, 5 and 7.
- 1 3. The polynucleotide of claim 1, wherein the fucosyltransferase
2 catalyzes the transfer of fucose to an N-acetylglucosamine residue, and wherein the nucleic
3 acid sequence has greater than 90% identity to a nucleotide sequence selected from the group
4 consisting of SEQ ID NO:1, 3, and 7.
- 1 4. The polynucleotide of claim 1, wherein the fucosyltransferase
2 catalyzes the transfer of fucose to a glucose residue, and wherein the nucleotide acid
3 sequence has greater than 90% identity to SEQ ID NO:5.
- 1 5. An isolated polynucleotide comprising a nucleic acid sequence,
2 wherein the nucleic acid sequence encodes a fucosyltransferase that catalyzes the transfer of a
3 fucose residue from a donor substrate to an acceptor substrate, and wherein the
4 fucosyltransferase comprises an amino acid selected from the group consisting of SEQ ID
5 NO:2, 4, 6, and 8.
- 1 6. The polynucleotide of claim 5, wherein the fucosyltransferase
2 comprises an amino acid tag.
- 1 7. An expression vector comprising the isolated polynucleotide of claim 1
2 or claim 5.
- 1 8. A host cell comprising the expression vector of claim 7.
- 1 9. A method of producing a fucosyltransferase protein, the method
2 comprising the step of culturing the host cell of claim 8 under conditions suitable for
3 expression of the fucosyltransferase protein.

1 10. An isolated polynucleotide comprising a nucleic acid sequence,
2 wherein the nucleic acid sequence consists of SEQ ID NO:11, wherein the nucleotide
3 sequence encodes a biologically active fucosyltransferase that catalyzes the transfer of a
4 fucose residue from a donor substrate to an acceptor substrate.

1 11. The polynucleotide of claim 10, wherein the fucosyltransferase
2 catalyzes the transfer of fucose to a glucose residue.

1 12. An isolated polynucleotide comprising a nucleic acid sequence,
2 wherein the nucleic acid sequence encodes a biologically active fucosyltransferase, and
3 wherein the fucosyltransferase comprises an amino acid that consists of SEQ ID NO:12.

1 13. An expression vector comprising the isolated polynucleotide of claim
2 10 or claim 12.

1 14. A host cell comprising the expression vector of claim 13.

1 15. A method of producing a fucosyltransferase protein, the method
2 comprising the step of culturing the host cell of claim 14 under conditions suitable for
3 expression of the fucosyltransferase protein.

1 16. A recombinant fucosyltransferase protein comprising a polypeptide has
2 greater than 90% identity to an amino acid sequence selected from the group consisting of
3 SEQ ID NO:2, 4, 6, and 8, wherein the fucosyltransferase catalyzes the transfer of a fucose
4 residue from a donor substrate to an acceptor substrate.

1 17. The recombinant fucosyltransferase of claim 16, further comprising an
2 amino acid tag.

1 18. The recombinant fucosyltransferase of claim 16, wherein the
2 polypeptide is selected from the group consisting of SEQ ID NO: 2, 4, 6, and 8.

1 19. The recombinant fucosyltransferase of claim 16, wherein the
2 fucosyltransferase catalyzes the transfer of fucose to an N-acetylglucosamine residue, and
3 wherein the polypeptide has greater than 90% identity to an amino acid sequence selected
4 from the group consisting of SEQ ID NO:2, 4, and 8.

1 20. The recombinant fucosyltransferase of claim 16, wherein the
2 fucosyltransferase catalyzes the transfer of fucose to a glucose residue, and wherein the
3 polypeptide has greater than 90% identity SEQ ID NO:6.

1 21. A recombinant fucosyltransferase protein comprising a polypeptide
2 consists of SEQ ID NO:12, wherein the fucosyltransferase catalyzes the transfer of a fucose
3 residue from a donor substrate to an acceptor substrate.

1 22. The recombinant fucosyltransferase of claim 21, wherein the
2 fucosyltransferase catalyzes the transfer of fucose to glucose.

1 23. A method of making a fucosylated oligosaccharide, the method
2 comprising:
3 contacting the recombinant fucosyltransferase of claim 16 with a mixture
4 comprising a donor substrate comprising a fucose residue, and an acceptor substrate
5 comprising a sugar or oligosaccharide, under conditions where the fusion protein catalyzes
6 the transfer of a fucose residue from the donor substrate to the acceptor substrate, thereby
7 producing a fucosylated oligosaccharide.

1 24. The method of claim 23, wherein the method further comprises a step
2 of purifying the fucosylated oligosaccharide.

1 25. The method of claim 23, wherein a donor substrate is GDP-fucose.

1 26. The method of claim 23, wherein the fucosyltransferase comprises an
2 amino acid tag.

1 27. The method of claim 23, wherein an acceptor substrate comprises a
2 member selected from N-acetylglucosamine and glucose.

1 28. The method of claim 23, wherein the acceptor substrate is Lacto-N-
2 neo-Tetraose (LNnT).

1 29. The method of claim 28, wherein the fucosylated oligosaccharide is
2 Lacto-N-Fucopentaose III (LNFP III).

1 30. The method of claim 23, wherein the mixture further comprises
2 lactose, a β -1,3-N-acetylglucosaminyltransferase, and a β -1,4-galactosyltransferase.

1 31. The method of claim 30, wherein the β -1,3-N-
2 acetylglucosaminyltransferase is a bacterial enzyme.

1 32. The method of claim 31, wherein the β -1,3-N-
2 acetylglucosaminyltransferase is from *Neisseria gonococcus*.

1 33. The method of claim 30, wherein the β -1,4-galactosyltransferase is a
2 bacterial enzyme.

1 34. The method of claim 33, wherein the β -1,4-galactosyltransferase is
2 from *Neisseria gonococcus*.

1 35. The method of claim 30, wherein the fucosylated oligosaccharide is
2 Lacto-N-Fucopentaose III (LNFP III).

1 36. A method for producing a fucosylated glycolipid, the method
2 comprising:
3 contacting the recombinant fucosyltransferase protein of claim 16 with a
4 mixture comprising a donor substrate comprising a fucose residue, and an acceptor substrate
5 on a glycolipid, under conditions where the fucosyltransferase catalyzes the transfer of the
6 fucose residue from a donor substrate to the acceptor substrate on the glycolipid, thereby
7 producing a fucosylated glycolipid.

FIGURE 1

Fucosyltransferase nucleotide sequence from strain 1182 FutB (SEQ ID NO:1)

atgttccaaccctattagacgcttatatagaaagcgcttcattgaaaaattacctctaaatctccccccccctaaaaatcgctg
tggcgaattggtggggagatgaagaggttgaagaatttaaaagaacattctttatttctcagtcagcattacacaatcaccct
ccacaaaacccaacgaacctccgatctcgtcttggcagtcctattggatcagccagaaaaatctatcctatcaaacgcaa
aaagagtgtttacaccggtgaaaacgaatcgctaatttcaacctcttgattacgccataggcttgatgaattggatttagagat
cgttatttaagaatgcctttatattatgatagactacaccataagccgagagcgtgaatgacaccacttcgcttacaactcaaac
ctgacagcctttatgctttaaaaaaacctcccatttttaagaaaaccacccaatttatgcgcagtagtgaacaatgagagcg
atcctttgaaaagaggggttgcgagttttagcgcagcaaccttaacgctcctaaaaggaatgctttctatgacgttttaattctata
gagccagttattgggggagggagcgtgaaaaacacttttagctataacattaaaaacaagagcgagttttaagccaatacaaat
tcaatctgtgtttgaaaactcacaaggctatggctatgtaactgaaaaaatcattgacgcttactttagccataaccattcctattattg
ggggagtcctagcgtggcacaagattttaaccctaagagtttgaatgtttgtgattttaagattttgatgaagcgattgatcatgt
gcgatacttgacacgcacccaaacgcttatttagacatgctttatgaaaacctttaaacaccttgatgggaaagcttactttac
caaatgtgagtttaaaaaaatcctagattttttaaacgatttagaaaacgacacgatttatcacgataacctttattttatcgt
gatttgaatgagccgttaatatctattgatgatgattgagggtaattatgatgattgagggtaattatgatgattgagggtaatta
tgatgattgagggtaattatgatgattgagggtaattatgatgattgagggtaattatgatgattgagggtaattatgatgatt
gagggtaattatgatgattgagggtaattatgatgattgagggtaattatgatgattgagggtaattatgagcggctcttaca
aaacgcctcgctttattagaactctctcaaacaccacttttaaatctatcgcaaagcttatcaaaaatccttaccttgttgcgtgc
ggcgagaaagttgattaaaaattgggttgtaa

Protein sequence from strain 1182 FutB (SEQ ID NO:2)

mfqplldayiesasiekitskspplkiavanwwgdeeeefkknilyfilsqhytitlhqnpnepsdlvfgspigsarkilsy
qnakrvfytgenespnfnlfdyaigfdeldfrdrylrmplyydrllhhaesvndttspyklkpdslyalkkpshhfknhpnl
cavvnnesdplkrgfasfvasnnpnapkrnafydvlnsiepvigggsvkntlgylnknkseflsqyknlc fensqgygyvte
kiidayfshtipiywgspsvaqdfnpksfvnvcdfkdfdeaidhvrylhthpnayldmlyenplntldgkayfyqnlsfkki
ldffktilendtiyhdnpfifyrdlneplisiddlrnyddlrnyddlrnyddlrnyddlrnyddlrnyddlrnyddlrnydd
lrnyddlrnyddlrnyddlrnyerllqnaspllelsqnttfkiyrkayqkslpllaarklikklgl*

FIGURE 2

Fucosyltransferase from strain 1111 FutA

Nucleotide coding sequence (SEQ ID NO:3)

atgttccaaccctattagatgcctttatagaaagcgctccattgaaaaatggcctctaaatctccccccctaaaaatcgctgtgg
cgaattggtggggagatgaagaaattaaaaaatttaaaagagcgttctttatttctcctaagccagcattacacaatcactttaca
ccgaaaccctgataaacctgcggacatcgtctttgtaaccccttgatcagccagaaaaatcttctctatcaaaacgcaaaaa
gggtgttttacaccggtgaaaatgaagtcctaacctcattgattacgccataggcttgatgaattggactttagagatcgt
tatttgagaatgcctttgtattatgcctatttgcattataaagccgagcttgtaatgacaccacttcgccttataaactccaacctgaca
gcctttatgctttaaaaaaacctcccatcattttaagaaaaaccaccccaatttgcgcgagtagtgaataatgagagtgatcctttg
aaaagaggggttgcgagctttgcaagcaaccctaacgcctcctagaaggaacgcttttatgaggctttaaacgctattgagcc
agttgctgggggagggagcgtgaaaaacactttaggctataatgtcaaaaacaagagcgagttttaagccaatacaaaaitcaat
ctgtgtttgaaaacactcaaggctatggctatgtaactgaaaagatcattgacgcttatttcagccataaccattcctatttattggggg
agtcccagcgtggcgaaaagatttaaccctaagagtttgaatgtccatgatttcaacaactttgatgaagcgattgactatatca
gatactgcacagcaccacaaacgcttatttagacatgcaactatgaaaaccctttaaacactattgatgggaaagcttactttacca
aaatttgagttttaaaaaaatcctagattttttaaacgattttgaaaacgacacgatctatcacgataaccctttcattttctatcgtg
atttgaatgagccttcagtatctattgatggtttgagggtaattatgatgatttgagggtaattatgatgatttgagggtaattatgat
gatttgagggtaattatgagcgcctttacaaaacgcctcgcctttattagaactctctcaaaacaccacttttaaatctatcgaaa
gcttatcaaaaatccttgcctttgtgcgtgccataaggagatgggttaaaaagtaa

Protein sequence (SEQ ID NO:4)

mfqplldafiesaplkkwplnlplkiavanwwgdeeikkfkksvlyfilsqhytitlhrnpdkpadivfgnplgsarkilsy
qnakrvfygenevnpnfnldfyagfdeldfrdrylrmplyyaylhykaelvndttspyklqpdsllyalkkpshhfknhpn
lcavvnnesdplkrfsvasnpnprnafyealnaiepvagggsvkntlgynvknkseflsqykfnlcfentqgygyvt
ekiidayfshtipiywgspsvakdfnpksfvnvhdfnnfdeaidyirylhthpnayldmhyenplntidgkayfyqnlsfk
kildffktilendtiyhdnpfifyrdlnepsvsidglrvnyddlrnyddlrnyddlrnyerllqnaspllelsqnttfkiyrka
yqkslpllrairrwvkk*

FIGURE 3

Strain 1218 FutB nucleotide sequence (SEQ ID NO:5)

atgttccaaccctattagacgcttatatagaaagcgttccattgaaaaaattacctctaaatctccccccccctaaaaatcgctg
tggcgaattggtggggagatgaagagggtgaagaattaaaaagaacattctttatttctcagtcagcattacacaatcacct
ccacaaaaccccaacgaacctccgatctcgtctttggcagtcctattggatcagccagaaaaatctatcctatcaaacgcaa
aaagagtgtttacaccggtgaaaacgaatgcctaatttcaacctctttgattacgccataggcttgatgaattggatttagagat
cgttatttaagaatgcctttatattatgatagactacaccataaagccgagagcgtgaatgacaccacttcgccttacaactcaaac
ctgacagcctttatgctttaaaaaaacctcccatcattttaaagaaaaccacccaatttatgcgtagtagtgaacaatgagagcg
atcctttgaaaagagggtttgtagcgagcaaccctaacgctcctaaaagggaatgctttctatgacgctttaattctata
gagccagttattgggggagggagcgtgaaaaacactttaggctataacattaaaaacaagagcaggttttaagccaatacaaat
tcaatctgtgtttgaaaactcacaaggctatggctatgtaactgaaaaaatcattgacgcttactttagccataaccattcctattattg
ggggagtcttagcgtggcacaagatttaaccctaagagtttgtgaatgtttgtattttaagattttgatgaagcgattgatcatgt
gcgatacttgacacgcacccaaacgcttatttagacatgctttatgaaaaccctttaacacccttgatgggaaagcttactttac
caaaatttgagtttaaaaaaaccttagatttttaaaacgatcttagaaaacgacacgatttatcacgataaccctttattttatcgt
gatttgaatgagccgttaatatctattgatgatttgagggtaattatgatgatttgagggtaattatgatgatttgagggtaattatga
tgatttgagggtaattatgatgatttgagggtaattatgatgatttgagggtaattatgatgatttgagggtaattatgatgattga
gggtaattatgatgatttgagggtaattgtgatgatttgagggtaattatgatgatttgagggtaattatgagcggctctacaaa
acgcctcgcctttattagaactctctcaaacaccacttttaaaatctatcgcaaagcttatcaaaatccttacctttgttgcgtgcgg
cgagaaagttgattaaaaaattgggtttgtaa

Predicted protein strain 1218 FutB (SEQ ID NO:6)

mfqplldayiesasiekitskspplkiavanwwgdeeeefkknilyfilsqhytitlhqnpnepsdlvfgspigsarkilsy
qnakrvfyngenepnfnlfdyaigfdeldfrdrylrmplyydrllhkaesvndttspyklkpdslyalkkpshhfkcnhpn
cavvnnesdplkrghasfvasnpnapkrnafydaalsiepvigggsvkntlgyniknkseflsqyfnlcfensqgygyvte
kiidayfshtipiywgspsvaquadfnpsfvnvcdfkdfdeaidhvrylhthpnayldmlyenplntldgkayfyqnlsfkk
ldffktilendtiyhndpfifyrdlneplisiddlrnyddlrnyddlrnyddlrnyddlrnyddlrnyddlrnyddlrnyddlr
nyddlrnyddlrnyddlrnyerllqnaspillelseqnttfkiyrkayqkspllraarklikklgl*

FIGURE 4

Fucosyltransferase strain 19C2 FutB nucleotide sequence (SEQ ID NO:7)

atgttccaaccctattagacgcttatatagacagcaccggttagatgaaaccgattataagccccattaaatatagccctagcg
aattggtggcctttggataaaagagaaagcaaagggttagaaaaaatttatcttacatttcatttaagtcagcattacacaatcgc
tctccaccgaaaccctgataaacctgcggacatcgttttggttaaccccttgatcagccagaaaaatcctatcctatcaaaacg
ctaaaagggtgttttacaccggtgaaaacgaagtcctaattcaacctctttgattacgccataggcttgatgaattggactttaga
gatcgttatttgagaatgcctttatattatgatagactacaccataaaagccgagagcgtgaatgacaccaccgcacctacaagatt
aaatctgacagcctttatgctttaaaaaagccctcccatcattttaagaaaaccaccacatttatgcgcgctaataataatgaga
tcgatcctttgaaaagagggttgcgagccttgcgcaagcaaccctaacgcccctataaggaaacgctttctatgaggctttaattc
tattgagccagtactgggggaggggagcgtgagaaacactttaggctataacgtcaaaaacaaaacgaattttgagccaatac
aagttcaatctgtgctttgaaaacactcaaggctatggctatgttactgaaaaaatcattgacgcttactcagccacaccattcctat
ttattgggggggagtccttagcgtggcgaaagatttaacccc

Strain 19C2 FutB protein sequence (SEQ ID NO:8)

mfqplldayidstrldetdykplnlalanwwpldkreskgfrkkfilhflsqhytialhrnpdkpadivfgnplgsarkilsy
qnakrvfytgenevpnfnlfdyaigfdeldfrdrylmplyydrllhbkhaesvndttapykiksdslyalkkpshhfkcnhph
lcalinneidplkrghasfvasnpnapirnafyealnsiepvttgggsrvntlgynvknkneflsqykfnlcfentqgygyvtek
iidayfshtipiywggvpsvakdfnp

FIGURE 5

Strain 915 FutA fucosyltransferase nucleotide coding sequence (SEQ ID NO:9)

atggcctctaaatctccccccctaaaaatcgctgtggcgaattggtggggagatgaagaaattaaaaattaaaaagagcgttct
ttattttatcctaagccagcattacacaatcactttacaccgaaaccctgataaacctgcggacatcgtctttgtaacccccttgat
cagccagaaaaatctatcctatcaaaacgcaaaaagggtgtttacaccggtgaaatgaagtcctaacttaacctcttgatta
cgccataggctt

Protein sequence from Strain 915 FutA (SEQ ID NO:10)

maskspplkiavanwwgdeeikkfkksvlyfilsqhytitlhrnpdkpdivfgnplgsarkilsyqnakrvfytgenevpn
fnlfdyai

FIGURE 6

Strain 26695 FutA fucosyltransferase nucleotide coding sequence (SEQ ID NO:11)

atgttccaaccctattagacgcctttatagaagcgcttccattgaaaaaatggcctctaaatctccccccccccctaaaaatc
gctgtggcgaattgggtgggagatgaagaaattaaagaattaaaaagagcggttctttatcttaagccaacgctacgcaatc
accctccacaaaacccaatgaatttcagatctagtttttagcaatcctcttgagcggttagaagattttatctatcaaacac
taaacgagtgtttacaccggtgaaaacgaatcacctaatttcaacctcttgattacgccataggctttgatgaattggattttaatga
tcgttatttgagaatgcctttgtattatgccatttgcactataaagccgagcttgtaatgacaccactgcgccctacaaactcaag
acaacagcctttatgctttaaaaaaacctctcatcttttaaagaaaaccaccctaatttgtgcgcagtagtgatgatgagagcg
atcttttaaaaagaggggttgcagttttagtcgagcaacgctaacgctcctatgaggaacgcttttatgacgctctaaattccata
gagccagttactgggggagggaagtgtgagaaacactttaggctataaggttggaacaaaagcgagttttaagccaatacaagt
tcaatctctgtttgaaaactcgcaagggttatggctatgaaccgaaaaaatccttgatgcgtattttagccataaccattcctatttattg
ggggagtcaccagcgtggcgaaagattttaaccctaaaagtgttgatgtcatgattcaacaactttgatgaagcgattgattat
atcaaatacctgcacacgcacccaaacgcttatttagacatgctctatgaaaaccctttaaacacccttgatgggaaagcttactttt
accaagatttgagttttaaaaaaatcctagattttttaaaacgattttagaaaacgatacgatttatcacaattctcaacatcttcatg
tgggagtacgatctgcataagccgttagtatccattgatgattgagggttaattatgatgatttgagggttaattatgaccggcttta
caaacgcttcgcctttattagaactctctcaaacaccacttttaaatctatcgcaaagcttatcaaaaatccttgctttgtgcgc
gcggtgagaaagttgggttaaaaattgggtttgtaa

Protein coding sequence Strain 26695 FutA (SEQ ID NO:12)

mfqplldafiesasiekmaskpppplkiavanwwgdeeikefkksvlyfilsqryaitlhqnpnefsdlvsnplgaarkil
syqntkrvfytgenespnfnldyaigfdeldfndrylrmplyyahlykaelvndttapyklkdnslyalkkpshhfkcnh
pnlcavvndesdllkrghasfvasnanapmrnafydaalnsieptgggsrvntlgkyvgnkseflsqyknfclfensqgygy
vtekildayfshtipiywgspsvakdfnpksfvnvhdnnfdeaidyikylhthpnayldmlyenplntldgkayfyqdlf
kkildffktilendtiyhkfstsfmweydlhkplvsiddlrnvnyddlrnvnydrllqnaspllelsqnttfkiyrkayqkslplrav
rklvkkgl*

FIGURE 7

19C2A fucosyltransferase nucleotide sequence (SEQ ID NO:13)

atgtccaacccttactagacgcctttatagaaagtgtccaatt

19C2A predicted protein sequence (SEQ ID NO:14)

mfqplldafiesapi

FIGURE 8

Protein sequence from strain 1182 FutB aligned with pfam00852, Glyco_transf_10,
Glycosyltransferase family 10

```
Query: 23 PPPLKIAVANWWGDEEVVEEFKKNILYFILSQHYTITLHQNPNEPSDLVFGS-PIGSARKI 81
Sbjct: 11 TVPLLLAIYTWWSLIEYKEWKKSPIYFIGSQAPQPPLR---ILLWTWPFNGNPLALSDCP 67

Query: 82 LSYQNAKRVPFYTGEN---ESPNNLF---DYAIGFDELDFRDRYLRMPLYDYDRLHHKAES 135
Sbjct: 68 LSYQNTARCRLTANRSPLESADAVLFHHRDLSKGFDPDLPSPRPFGQPWVWASMESPSNS 127

Query: 136 -VND'TTSPYKLKPDLSYALKKPSHHFKENHPNLCVVNNEDEPLKRGFASFVASNP-AP 193
Sbjct: 128 GLNDLRDGYFNWTLSTYRADSDAFHPYGYLEPRLSQVVNAPLLSAKRKGAAWVVSNCNTRS 187

Query: 194 KRNAFYDVLNSIEPVIIGGGSVKNTLGYNINKNSEFLSQYKFNLCFENSQGYGYVTEKIID 253
Sbjct: 188 KRERFYKQLNKLQVDVGGRVANPLPLKVGCLVETLSQYKFYLAFENSQHYDYVTEKLWK 247

Query: 254 -AYFSHTIPIYWGSPSVAQDFNP-KSFVNVCDFKDFDEAIDHVRYLHTHPNAYL 305
Sbjct: 248 NALQAGTIPVVLGPRAVYEDFVPPKSFIVDDFKSPKELADYLLLYLDTNPTAYS 301
```

FIGURE 9

Fucosyltransferase from strain 1111 FutA aligned with pfam00852, Glyco_transf_10,
Glycosyltransferase family 10

```

Query:  27  IAVANWWGDEEIKKFKKSVLYFILSQHYTITLHRNPDKPADIVFG-NPLGSARKILSYQN  85
Sbjct:  16  LAIYTWWSLIEYKEWKKSPIYFIGSQAPQPPLR---ILLWTWPFNGNPLALSDCPLSYQN  72

Query:  86  AKRVFYTGGEN---EVPNFNLF---DYAIGFDELDLDRDRLRMPLYAYLHYKAEL-VNDT  138
Sbjct:  73  TARCRLTANRSPLESADAVLFHHRDLSKGFDPDLPPSPRPPGQPWVWASMESPSNSGLNDL  132

Query:  139  TSPYKLQPDLSLYALKKPSHHFKENHPNLCAVVNNESDPLKRGFASFVASNPN-APRRNAF  197
Sbjct:  133  RDGYFNWTLSTYRADSDAFHPYGYLEPRLSQVVNAPLLSAKRKGAANVVSNCTRKRERF  192

Query:  198  YEALNAIEPVAGGGSVKNTLGYNVKNKSEFLSQYKFNLCFENTQGYGYVTEKIID-AYFS  256
Sbjct:  193  YKQLNKHLQVDVGGRVANPLPLKVGCLVETLSQYKFYLAFENSQHYDYVTEKLWKNALQA  252

Query:  257  HTIPIYWGSPSVAKDFNP-KSFVNVHDFNNFDEAIDYIRYLHTHPNAYLDMHYENPLNTI  315
Sbjct:  253  GTIPVVLGPRAVYEDFVPPKSFHVDDFKSPKELADYLLYLDTNPTAYS-----  301

Query:  316  DGKAYFYQNLSFKKILDFFKTILENDTIYHDNPFIFYRDLNEPSVSIDGLRVNYDDLVRN  375
Sbjct:  302  -----EYFEWRYDLRVRLFSDALR-----  321

Query:  376  YDDLVRVNYDDLVRVNYERLLQNASPLLELSQNTTFKIYRKAYQ  417
Sbjct:  322  -----YDEGFCRVCRLQNAPD-----RYKTYPNIAKWFQ  351

```

FIGURE 10

Protein sequence from strain 1218 FutB aligned with pfam00852, Glyco_transf_10,
Glycosyltransferase family 10

```

Query:  23  PPPLKIAVANWWGDEEVEEFKKNILYFILSQHYTITLHQNPNEPSDLVFGS-PIGSARKI   81
Sbjct:  11  TVPLLLLAIYTWWSLIEYKEWKKSPIYFIGSQAPQPPLR---ILLWTWPFNGNPLALSDCP . 67

Query:  82  LSYQNAKRVFYTGEN---ESPNNLF---DYAIGFDELDLFRDRLRMPLYDRLHHAES   135
Sbjct:  68  LSYQNTARCRLTANRSPLESADAVLFHHRDL SKGFDDLPPSPRPPGQPWVWASMESPSNS   127

Query:  136 -VNDTTSPIYKLKPDLSLYALKKPSHHFKENHPNLCAVVNNESDPLKRGFASFVASNP-AP   193
Sbjct:  128 GLNDLRDGYFNWTL SYRADSDAFHPYGYLEPRLSQVVNAPLLSAKRKGAAWVVSNCNTRS   187

Query:  194 KRNAFYDALNSIEPVIGGGSVKNTLGYNIKNKSEFLSQYKFNLCFENSQGYGYVTEKIID   253
Sbjct:  188 KRERFYKQLNKKHLQVDVGGRVANPLPLKVGCLVETLSQYKFYLA FENSQHYDYVTEKLWK   247

Query:  254 -AYFSHTIPIYWGPSVAQDFNP-KSFVNVCDFKDFDEAIDHVRYLHHPNAYLDMLYEN   311
Sbjct:  248 NALQAGTIPVVLGPRAVYEDFVPPKSFIVHDDFKSPKELADYLLYLDTNPTAYS-----   301

Query:  312 PLNTLDGKAYFYQNL SFKKILDFFKTILENDTIYHDNPFIFYRDLNEPLISIDDLRVNYD   371
Sbjct:  302 -----EYFEWRYDLRVRLFSWDALR--YD   323

Query:  372 DLRVNYDDL RVNYDDL RVNYDDL RVNYD   399
Sbjct:  324 EGFCRVCRLLQNAFDRYKTYPNIAKWFQ   351

```

FIGURE 11

Protein sequence from strain 19C2 FutB aligned with pfam00852, Glyco_transf_10,
Glycosyltransferase family 10

Query:	22	PPLNIALANWWPLDKRESKGFRKKFILHFILSQHYTIALHRNPDKPADIVFG-NPLGSAR	80
Sbjct:	12	VPLLLAIYTWWSL--IEYKEW-KKSPIYFIGSQAPQPPLR---ILLWTWPFNGNPLALSD	65
Query:	81	KILSYQNAKRVFYTGGEN---EVPNFNLF---DYAIGFDELDFRDRYLRMPLYDRLHHKA	134
Sbjct:	66	CPLSYQNTARCRLTANRSPLESADAVLFHHRDLKGFPDLPSPRPPGQPWVWASMESPS	125
Query:	135	ES-VNDTTAPYKIKSDSLYALKKPSHHFKENHPHLCALINNEIDPLKRGFASFVASNPN-	192
Sbjct:	126	NSGLNDLRDGYFNWTLNLRADSDAFHPYGYLEPRLSQVVNAPLLSAKRKGAAWVVSNCNT	185
Query:	193	APIRNAFYEALNSIEPVTGGGSVRNTLGYNVKNKNEFLSQYKFNLCFENTQGYGYVTEKI	252
Sbjct:	186	RSKRERFYKQLNKHQLQVDVGGRVANPLPLKVGCLVETLSQYKFYLAFENSQHYDYVTEKL	245
Query:	253	ID-AYFSHTIPIYWGGVPSVAKDFNP	277
Sbjct:	246	WKNALQAGTIPVVLGP-RAVYEDFVP	270

FIGURE 12

	1	50
1111FutA.pep	(1) MFQPLLDAFIESAPLKKWPLN--LPPLKIAVANWWGDEETKK---FKKSV	
19C2A.pep	(1) MFQPLLDAFIESAPI-----	
915A.pepneose	(1) -----MASK-SPPLKIAVANWWGDEETKK---FKKSV	
26695A.pep	(1) MFQPLLDAFIESASIEKMASKSPPPPLKIAVANWWGDEETKE---FKKSV	
1182B.pep	(1) MFQPLLDAYIESASIEKITSKS-PPPLKIAVANWWGDEEVEE---FKKNT	
1218B.pep	(1) MFQPLLDAYIESASIEKITSKS-PPPLKIAVANWWGDEEVEE---FKKNT	
ORF19C2B.pep	(1) MFQPLLDAYIDSTRIDETDYK---PPLNIALANWWPLDKRESKGFRRKKFI	
Consensus	(1) MFQPLLDAFIESA IEK SK PPLKIAVANWWGDEEI FKK I	
	51	100
1111FutA.pep	(46) LYFILSQHYTITLHRNPDKPADIVFGNPLGSARKILSYQNAKRVFYTG	
19C2A.pep	(16) -----	
915A.pepneose	(29) LYFILSQHYTITLHRNPDKPADIVFGNPLGSARKILSYQNAKRVFYTG	
26695A.pep	(48) LYFILSQRYAITLHQNPNEPSDLVFSNPLGAARKILSYQNTKRVFYTG	
1182B.pep	(47) LYFILSQHYTITLHQNPNEPSDLVFGSPIGSARKILSYQNAKRVFYTG	
1218B.pep	(47) LYFILSQHYTITLHQNPNEPSDLVFGSPIGSARKILSYQNAKRVFYTG	
ORF19C2B.pep	(48) LYFILSQHYTIALHRNPDKPADIVFGNPLGSARKILSYQNAKRVFYTG	
Consensus	(51) LYFILSQHYTITLH NP PADIVFGNPLGSARKILSYQNAKRVFYTG	
	101	150
1111FutA.pep	(96) EVPNFNLFDYAIGFDELDFRDRYLRMPLYAYLHYKAELVNDTTSPYKLQ	
19C2A.pep	(16) -----	
915A.pepneose	(79) EVPNFNLFDYAIGF-----	
26695A.pep	(98) ESPNFNLFDYAIGFDELDFNDRYLRMPLYYAHLHYKAELVNDTTAPYKLK	
1182B.pep	(97) ESPNFNLFDYAIGFDELDFRDRYLRMPLYYDRLHHKAESVNDTTSPYKLK	
1218B.pep	(97) ESPNFNLFDYAIGFDELDFRDRYLRMPLYYDRLHHKAESVNDTTSPYKLK	
ORF19C2B.pep	(98) EVPNFNLFDYAIGFDELDFRDRYLRMPLYYDRLHHKAESVNDTTAPYKLK	
Consensus	(101) E PNFNLFDYAIGFDELDFRDRYLRMPLY Y LHHKAE VNDTTSPYKLK	
	151	200
1111FutA.pep	(146) PDSLYALKKPSHHFKENHPNLCVVNNESDPLKRGFASFVASNPNA PRRN	
19C2A.pep	(16) -----	
915A.pepneose	(93) -----	
26695A.pep	(148) DNSLYALKKPSHHFKENHPNLCVVNNESDLLKRGFASFVASNANAPMRN	
1182B.pep	(147) PDSLYALKKPSHHFKENHPNLCVVNNESDPLKRGFASFVASNPNA PKRN	
1218B.pep	(147) PDSLYALKKPSHHFKENHPNLCVVNNESDPLKRGFASFVASNPNA PKRN	
ORF19C2B.pep	(148) SDSLYALKKPSHHFKENHPNLCALINNEIDPLKRGFASFVASNPNA PIRN	
Consensus	(151) DSLYALKKPSHHFKENHPNLCVVNNESDPLKRGFASFVASNPNA P RN	
	201	250
1111FutA.pep	(196) AFYEALNAIEPVAGGGSVKNTLGYNVKNKSEFLSQYKFNL CFENTQGYGY	
19C2A.pep	(16) -----	
915A.pepneose	(93) -----	
26695A.pep	(198) AFYDALNSIEPVTGGGSVRNTLGYSKVGKNKSEFLSQYKFNL CFENSQGYGY	
1182B.pep	(197) AFYDVLNSIEPVIGGGSVKNTLGYNIKNKSEFLSQYKFNL CFENSQGYGY	
1218B.pep	(197) AFYDALNSIEPVIGGGSVKNTLGYNIKNKSEFLSQYKFNL CFENSQGYGY	
ORF19C2B.pep	(198) AFYEALNSIEPVTGGGSVRNTLGYNVKNKNEFLSQYKFNL CFENTQGYGY	
Consensus	(201) AFYDALNSIEPV GGGSVKNTLGYNVKNKSEFLSQYKFNL CFENSQGYGY	

	1		50
1111FutA	(1)	ATGTTCCAACCCCTATTAGATGCCTTTATAGAAAGCGCT-CCATTGAAAA	
915A.cod (MWG)	(1)	ATGTTCCAACCCCTATTAGATGCCTTTATAGAAAGCGCTTCCATTGAAAA	
19C2FutA.cod	(1)	ATGTTCCAACCCCTTACTAGACGCCTTTATAGAAAGTGCTCCAATT-----	
26695A.cod	(1)	ATGTTCCAACCCCTATTAGACGCCTTTATAGAAAGCGCTTCCATTGAAAA	
1182B	(1)	ATGTTCCAACCCCTATTAGACGCCTTATATAGAAAGCGCTTCCATTGAAAA	
1218B.nuc	(1)	ATGTTCCAACCCCTATTAGACGCCTTATATAGAAAGCGCTTCCATTGAAAA	
ORF19C2B	(1)	ATGTTCCAACCCCTATTAGACGCCTTATATAGACGACACCCGTTTAGATGA	
Consensus	(1)	ATGTTCCAACCCCTATTAGACGCCTTTATAGAAAGCGCTTCCATTGAAAA	
	51		100
1111FutA	(50)	AATGGCCTCTAAATCTCCCCCCCC-----TAAAAATCGCTGTGGCGAATT	
915A.cod (MWG)	(51)	AATGGCCTCTAAATCTCCCCCCCC-----TAAAAATCGCTGTGGCGAATT	
19C2FutA.cod	(46)	-----	
26695A.cod	(51)	AATGGCCTCTAAATCTCCCCCCCCCCCCCTAAAAATCGCTGTGGCGAATT	
1182B	(51)	AATTACCTCTAAATCTCCCCCCCCCCCC-----TAAAAATCGCTGTGGCGAATT	
1218B.nuc	(51)	AATTACCTCTAAATCTCCCCCCCCCCCC-----TAAAAATCGCTGTGGCGAATT	
ORF19C2B	(51)	AACCGATTATAA-----GCCCCCAT--TAAATATAGCCCTAGCGAATT	
Consensus	(51)	AAT GCCTCTAAATCTCCCCCCCC TAAAAATCGCTGTGGCGAATT	
	101		150
1111FutA	(95)	GGTGG-----GGAGATGA-AGAAATTAAAAAATTTAAAAAGAGCGTTCTT	
915A.cod (MWG)	(95)	GGTGG-----GGAGATGA-AGAAATTAAAAAATTTAAAAAGAGCGTTCTT	
19C2FutA.cod	(46)	-----	
26695A.cod	(101)	GGTGG-----GGAGATGA-AGAAATTAAAGAATTTAAAAAGAGCGTTCTT	
1182B	(98)	GGTGG-----GGAGATGA-AGAGGTTGAAGAATTTAAAAAGAACATTCTT	
1218B.nuc	(98)	GGTGG-----GGAGATGA-AGAGGTTGAAGAATTTAAAAAGAACATTCTT	
ORF19C2B	(92)	GGTGGCCTTTGGATAAAAGAGAAAGCAAAGGGTTTAGAAAAAATTTATC	
Consensus	(101)	GGTGG GGAGATGA AGAAATTAAAGAATTTAAAAAGA C TTCTT	
	151		200
1111FutA	(139)	T---ATTTTATCCTAAGCCAGCATTACACAATCACTTTACACCGAAACCC	
915A.cod (MWG)	(139)	T---ATTTTATCCTAAGCCAGCATTACACAATCACTTTACACCGAAACCC	
19C2FutA.cod	(46)	-----	
26695A.cod	(145)	T---ATTTTATCCTAAGCCAACGCTACGCAATCACCTCGACCAAAACCC	
1182B	(142)	T---ATTTTATCTCAGTCAGCATTACACAATCACCTCCACCAAAACCC	
1218B.nuc	(142)	T---ATTTTATCTCAGTCAGCATTACACAATCACCTCCACCAAAACCC	
ORF19C2B	(142)	TTACATTTTATTTTAAGTCAGCATTACACAATCGCTCTCCACCGAAACCC	
Consensus	(151)	T ATTTTAT CTAAG CAGCATTACACAATCAC CTCCACC AAACCC	
	201		250
1111FutA	(186)	TGATAAACCTGCGGACATCGTCTTTGGTAACCCCTTTGGATCAGCCAGAA	
915A.cod (MWG)	(186)	TGATAAACCTGCGGACATCGTCTTTGGTAACCCCTTTGGATCAGCCAGAA	
19C2FutA.cod	(46)	-----	
26695A.cod	(192)	CAATGAATTTTCAGATCTAGTTTTTAGCAATCCTCTTTGGAGCGGCTAGAA	
1182B	(189)	CAACGAACCCCTCCGATCTCGTCTTTGGCAGTCCTATTGGATCAGCCAGAA	
1218B.nuc	(189)	CAACGAACCCCTCCGATCTCGTCTTTGGCAGTCCTATTGGATCAGCCAGAA	
ORF19C2B	(192)	TGATAAACCTGCGGACATCGTTTTTTGGTAACCCCTTTGGATCAGCCAGAA	
Consensus	(201)	AT AACCT C GA TCGTCTTTGG AA CC CTTGGATCAGCCAGAA	

	251	300
1111FutA	(236)	AAATCTTATCCTATCAAAAACGCAAAAAGGGTGTTTTACACCGGTGAAAAT
915A.cod (MWG)	(236)	AAATCTTATCCTATCAAAAACGCAAAAAGGGTGTTTTACACCGGTGAAAAT
19C2FutA.cod	(46)	-----
26695A.cod	(242)	AGATTTTATCCTTATCAAAAACACTAAACGAGTGTTTTACACCGGTGAAAAC
1182B	(239)	AAATCTTATCCTATCAAAAACGCAAAAAGAGTGTTTTACACCGGTGAAAAC
1218B.nuc	(239)	AAATCTTATCCTATCAAAAACGCAAAAAGAGTGTTTTACACCGGTGAAAAC
ORF19C2B	(242)	AAATCCTATCCTATCAAAAACGCTAAAAGGGTGTTTTACACCGGTGAAAAC
Consensus	(251)	AAATCTTATCCTATCAAAAACGCAAAAAG GTGTTTTACACCGGTGAAAAC
	301	350
1111FutA	(286)	GAAGTCCCTAACTTCAACCTCTTTGATTACGCCATAGGCTTT-GATGAAT
915A.cod (MWG)	(286)	GAAGTCCCTAACTTCAACCTCTTTGATTACGCCATAGGCTTTTGATGA--
19C2FutA.cod	(46)	-----
26695A.cod	(292)	GAATCACCTAATTTCAACCTCTTTGATTACGCCATAGGCTTT-GATGAAT
1182B	(289)	GAATCGCCTAATTTCAACCTCTTTGATTACGCCATAGGCTTT-GATGAAT
1218B.nuc	(289)	GAATCGCCTAATTTCAACCTCTTTGATTACGCCATAGGCTTT-GATGAAT
ORF19C2B	(292)	GAAGTCCCTAATTTCAACCTCTTTGATTACGCCATAGGCTTT-GATGAAT
Consensus	(301)	GAA CCTAATTTCAACCTCTTTGATTACGCCATAGGCTTT GATGAAT
	351	400
1111FutA	(335)	TGGACTTTAGAGATCGTTATTTGAGAATGCCTTTGTATTATGCCTATTTG
915A.cod (MWG)	(334)	-----
19C2FutA.cod	(46)	-----
26695A.cod	(341)	TGGATTTTAATGATCGTTATTTGAGAATGCCTTTGTATTATGCCCATTTG
1182B	(338)	TGGATTTTAGAGATCGTTATTTAAGAATGCCTTTATATTATGATAGACTA
1218B.nuc	(338)	TGGATTTTAGAGATCGTTATTTAAGAATGCCTTTATATTATGATAGACTA
ORF19C2B	(341)	TGGACTTTAGAGATCGTTATTTGAGAATGCCTTTATATTATGATAGACTA
Consensus	(351)	TGGA TTTAGAGATCGTTATTT AGAATGCCTTT TATTATG T
	401	450
1111FutA	(385)	CATTATAAAGCCGAGCTTGTTAATGACACCACTTCGCCTTATAAACTCCA
915A.cod (MWG)	(334)	-----
19C2FutA.cod	(46)	-----
26695A.cod	(391)	CACATATAAAGCCGAGCTTGTTAATGACACCACTTCGCCTTACAAACTCAA
1182B	(388)	CACCATAAAGCCGAGAGCGTGAATGACACCACTTCGCCTTACAAACTCAA
1218B.nuc	(388)	CACCATAAAGCCGAGAGCGTGAATGACACCACTTCGCCTTACAAACTCAA
ORF19C2B	(391)	CACCATAAAGCCGAGAGCGTGAATGACACCACTTCGCCTTACAAAGATTAA
Consensus	(401)	CAC ATAAAGCCGAG GT AATGACACCACT CGCCTTACAAACTCAA
	451	500
1111FutA	(435)	ACCTGACAGCCTTTATGCTTTAAAAAAACCCTCCCATCATTTTAAAGAAA
915A.cod (MWG)	(334)	-----
19C2FutA.cod	(46)	-----
26695A.cod	(441)	AGACAACAGCCTTTATGCTTTAAAAAAACCCTCTCATCATTTTAAAGAAA
1182B	(438)	ACCTGACAGCCTTTATGCTTTAAAAAAACCCTCCCATCATTTTAAAGAAA
1218B.nuc	(438)	ACCTGACAGCCTTTATGCTTTAAAAAAACCCTCCCATCATTTTAAAGAAA
ORF19C2B	(441)	ATCTGACAGCCTTTATGCTTTAAAAAAGCCCTCCCATCATTTTAAAGAAA
Consensus	(451)	A CTGACAGCCTTTATGCTTTAAAAAAACCCTCCCATCATTTTAAAGAAA

FIGURE 13 (cont'd)

	501	550
1111FutA	(485) ACCACCCCAATTTGTGCGCAGTAGTGAATAATGAGAGTGATCCTTTTGAAA	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(491) ACCACCCCTAATTTGTGCGCAGTAGTGAATGATGAGAGCGATCTTTTAAAA	
1182B	(488) ACCACCCCAATTTATGCGCAGTAGTGAACAATGAGAGCGATCCTTTTGAAA	
1218B.nuc	(488) ACCACCCCAATTTATGCGCAGTAGTGAACAATGAGAGCGATCCTTTTGAAA	
ORF19C2B	(491) ACCACCCACATTTATGCGCGCTAATCAATAATGAGATCGATCCTTTTGAAA	
Consensus	(501) ACCACCC AATTT TCGCAGTAGTGAA AATGAGAGCGATCCTTTTGAAA	
	551	600
1111FutA	(535) AGAGGGTTTGTGCGAGCTTTGTGCGCAAGCAACCCTAACGCTCCTAGAAAGGAA	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(541) AGAGGGTTTGTCCAGTTTGTAGCGAGCAACGCTAACGCTCCTATGAGGAA	
1182B	(538) AGAGGGTTTGTGCGAGTTTGTAGCGAGCAACCCTAACGCTCCTAAAAGGAA	
1218B.nuc	(538) AGAGGGTTTGTGCGAGTTTGTAGCGAGCAACCCTAACGCTCCTAAAAGGAA	
ORF19C2B	(541) AGAGGGTTTGTGCGAGCTTTGTGCGCAAGCAACCCTAACGCCCCATATAAGGAA	
Consensus	(551) AGAGGGTTTGTGCGAG TTTGT GC AGCAACCCTAACGCTCCTA AAGGAA	
	601	650
1111FutA	(585) CGCTTTTTTATGAGGCTTTTAAACGCTATTGAGCCAGTTGCTGGGGGAGGGA	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(591) CGCTTTTTTATGACGCTCTAAATTCCATAGAGCCAGTTACTGGGGGAGGAA	
1182B	(588) TGCTTTCTATGACGTTTTTAAATTCTATAGAGCCAGTTATTGGGGGAGGGA	
1218B.nuc	(588) TGCTTTCTATGACGTTTTTAAATTCTATAGAGCCAGTTATTGGGGGAGGGA	
ORF19C2B	(591) CGCTTTCTATGAGGCTTTTAAATTCTATTGAGCCAGTTACTGGGGGAGGGA	
Consensus	(601) GCTTT TATGA GCTTTAAATTCTAT GAGCCAGTTA TGGGGGAGGGA	
	651	700
1111FutA	(635) GCGTGAAAAACACTTTAGGCTATAATGTCAAAAACAAGAGCGAGTTTTTTA	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(641) GTGTGAGAAACACTTTAGGCTATAAGGTTGGAAACAAAAGCGAGTTTTTTA	
1182B	(638) GCGTGAAAAACACTTTAGGCTATAACATTAAAAACAAGAGCGAGTTTTTTA	
1218B.nuc	(638) GCGTGAAAAACACTTTAGGCTATAACATTAAAAACAAGAGCGAGTTTTTTA	
ORF19C2B	(641) GCGTGAGAAACACTTTAGGCTATAACGTCAAAAACAAAACGAATTTTTG	
Consensus	(651) GCGTGA AAACACTTTAGGCTATAA T AAAAACAA AGCGAGTTTTTTA	
	701	750
1111FutA	(685) AGCCAATACAAATTCAATCTGTGTTTTGAAAACACTCAAGGCTATGGCTA	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(691) AGCCAATACAAAGTTCAATCTGTGTTTTGAAAACCTCGCAAGGTTATGGCTA	
1182B	(688) AGCCAATACAAATTCAATCTGTGTTTTGAAAACCTCACAAGGCTATGGCTA	
1218B.nuc	(688) AGCCAATACAAATTCAATCTGTGTTTTGAAAACCTCACAAGGCTATGGCTA	
ORF19C2B	(691) AGCCAATACAAAGTTCAATCTGTGCTTTGAAAACACTCAAGGCTATGGCTA	
Consensus	(701) AGCCAATACAA TTCAATCTGTGTTTTGAAAAC C CAAGGCTATGGCTA	

	751	800
1111FutA	(735) TGTAAGTGAAGATCATTGACGCTTATTTAGCCATACCATTTCCTATTT	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(741) TGTAACCGAAAAAATCCTTGATGCGTATTTAGCCATACCATTTCCTATTT	
1182B	(738) TGTAAGTGAAGATCATTGACGCTTATTTAGCCATACCATTTCCTATTT	
1218B.nuc	(738) TGTAAGTGAAGATCATTGACGCTTATTTAGCCATACCATTTCCTATTT	
ORF19C2B	(741) TGTTACTGAAAAAATCATTGACGCTTATTTAGCCATACCATTTCCTATTT	
Consensus	(751) TGTAAGTGAAGATCATTGACGCTTA TT AGCCATACCATTTCCTATTT	
	801	850
1111FutA	(785) ATTGGGGG--AGTCC-CAGCGTGGCGAAAGATTTTAACCCTAAGAGTTTT	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(791) ATTGGGGG--AGTCC-CAGCGTGGCGAAAGATTTTAACCCTAAGAGTTTT	
1182B	(788) ATTGGGGG--AGTCC-TAGCGTGGCGACAAGATTTTAACCCTAAGAGTTTT	
1218B.nuc	(788) ATTGGGGG--AGTCC-TAGCGTGGCGACAAGATTTTAACCCTAAGAGTTTT	
ORF19C2B	(791) ATTGGGGGGGAGTCCCTAGCGTGGCGAAAGATTTTAACCCC-----	
Consensus	(801) ATTGGGGG AGTCC AGCGTGGC AAGATTTTAACCCTAA AGTTTT	
	851	900
1111FutA	(832) GTGAATGTCCATGATTTCAACAACCTTTGATGAAGCGATTGACTATATCAG	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(838) GTGAATGTGCATGATTTCAACAACCTTTGATGAAGCGATTGATTATATCAA	
1182B	(835) GTGAATGTTTGTGATTTTAAAGATTTTGTGAAGCGATTGATCATGTGCG	
1218B.nuc	(835) GTGAATGTTTGTGATTTTAAAGATTTTGTGAAGCGATTGATCATGTGCG	
ORF19C2B	(832) -----	
Consensus	(851) GTGAATGT TGATTT AA A TTTGATGAAGCGATTGA AT T	
	901	950
1111FutA	(882) ATACTTGCACACGCACCCAAACGCTTATTTAGACATGCATATGAAAACC	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(888) ATACCTGCACACGCACCCAAACGCTTATTTAGACATGCTCTATGAAAACC	
1182B	(885) ATACTTGCACACGCACCCAAACGCTTATTTAGACATGCTTTATGAAAACC	
1218B.nuc	(885) ATACTTGCACACGCACCCAAACGCTTATTTAGACATGCTTTATGAAAACC	
ORF19C2B	(832) -----	
Consensus	(901) ATAC TGCACACGCACCCAAACGCTTATTTAGACATGC TATGAAAACC	
	951	1000
1111FutA	(932) CTTTAAACACTATTGATGGGAAAGCTTACTTTTACCAAATTTGAGTTTT	
915A.cod (MWG)	(334) -----	
19C2FutA.cod	(46) -----	
26695A.cod	(938) CTTTAAACACCCTTGATGGGAAAGCTTACTTTTACCAAATTTGAGTTTT	
1182B	(935) CTTTAAACACCCTTGATGGGAAAGCTTACTTTTACCAAATTTGAGTTTT	
1218B.nuc	(935) CTTTAAACACCCTTGATGGGAAAGCTTACTTTTACCAAATTTGAGTTTT	
ORF19C2B	(832) -----	
Consensus	(951) CTTTAAACAC TTGATGGGAAAGCTTACTTTTACCAA ATTTGAGTTTT	

FIGURE 13 (cont'd)

		1001	1050
1111FutA	(982)	AAAAAAATCCTAGATTTTTTTTAAAAACGATTTTAGAAAACGACACGATCTA	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(988)	AAAAAAATCCTAGATTTTTTTTAAAAACGATTTTAGAAAACGATACGATTTA	
1182B	(985)	AAAAAAATCCTAGATTTTTTTTAAAAACGATTTTAGAAAACGACACGATTTA	
1218B.nuc	(985)	AAAAAAATCCTAGATTTTTTTTAAAAACGATCTTAGAAAACGACACGATTTA	
ORF19C2B	(832)	-----	
Consensus	(1001)	AAAAAAATCCTAGATTTTTTTTAAAAACGAT TTAGAAAACGA ACGAT TA	
		1051	1100
1111FutA	(1032)	TCACGATAACCC-----TTTCATTTTCTATCGTGATTGAATGAGCCGT	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1038)	TCACAAATTCTCAACATCTTTTCATGTGGGAGTACGATCTGCATAAGCCGT	
1182B	(1035)	TCACGATAACCC-----TTTTATTTTTCATCGTGATTGAATGAGCCGT	
1218B.nuc	(1035)	TCACGATAACCC-----TTTTATTTTTCATCGTGATTGAATGAGCCGT	
ORF19C2B	(832)	-----	
Consensus	(1051)	TCAC A C C TTT AT T A GAT TG AT AGCC T	
		1101	1150
1111FutA	(1076)	CAGTATCTATTGATGGT---TTGAGGGTTAATTATGATGATTTGAGGGTT	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1088)	TAGTATCCATTGATGAT---TTGAGGGTTAATTATGATGATTTGAGGGTT	
1182B	(1079)	TAATATCTATTGATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTT	
1218B.nuc	(1079)	TAATATCTATTGATGAT---TTGAGGGTTAATTATGATGATTTGAGGGTT	
ORF19C2B	(832)	-----	
Consensus	(1101)	A TATC ATTGATG T TTGAGGGTTAATTATGATGATTTGAGGGTT	
		1151	1200
1111FutA	(1123)	AATTATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGA	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1135)	AATTATGACCGGCTTTTACAAAACGCTTCGCCTTTATTAGAACTCTCTCA	
1182B	(1129)	AATTATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGA	
1218B.nuc	(1126)	AATTATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGA	
ORF19C2B	(832)	-----	
Consensus	(1151)	AATTATGA T AA T TTT G T T A	
		1201	1250
1111FutA	(1173)	GCGCCTTTTACAAAACGCCTCGCCTTTATTAGAACTCTCTCAAAACACCA	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1185)	AAACACCACTTTTAAATCTATCGCAAAGCTTATCAAAAATCCTTGCCCTT	
1182B	(1179)	TGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGATGATTTGA	
1218B.nuc	(1176)	TGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGATGATTTGA	
ORF19C2B	(832)	-----	
Consensus	(1201)	AA	

FIGURE 13 (cont'd)

		1251	1300
1111FutA	(1223)	CTTTTAAAAATCTATCGCAAAGCTTATCAAAAATCCTTGCCTTTGTTGCGT	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1235)	TGTTGCGCGCGGTGAGAAAGTTGGTTAAAAAATTGGGTTTGTA-----	
1182B	(1229)	GGGTTAATTATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAAT	
1218B.nuc	(1226)	GGGTTAATTATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAAT	
ORF19C2B	(832)	-----	
Consensus	(1251)	T	T A A
		1301	1350
1111FutA	(1273)	GCCATAAGGAGATGGGTTAAAAAGTAA-----	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1279)	-----	
1182B	(1279)	TATGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGAGCG	
1218B.nuc	(1276)	TGTGATGATTTGAGGGTTAATTATGATGATTTGAGGGTTAATTATGAGCG	
ORF19C2B	(832)	-----	
Consensus	(1301)		
		1351	1400
1111FutA	(1300)	-----	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1279)	-----	
1182B	(1329)	GCTCTTACAAAACGCCTCGCCTTTATTAGAACTCTCTCAAAACACCACTT	
1218B.nuc	(1326)	GCTCTTACAAAACGCCTCGCCTTTATTAGAACTCTCTCAAAACACCACTT	
ORF19C2B	(832)	-----	
Consensus	(1351)		
		1401	1450
1111FutA	(1300)	-----	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1279)	-----	
1182B	(1379)	TTAAAATCTATCGCAAAGCTTATCAAAAATCCTTACCTTTGTTGCGTGCG	
1218B.nuc	(1376)	TTAAAATCTATCGCAAAGCTTATCAAAAATCCTTACCTTTGTTGCGTGCG	
ORF19C2B	(832)	-----	
Consensus	(1401)		
		1451	1483
1111FutA	(1300)	-----	
915A.cod (MWG)	(334)	-----	
19C2FutA.cod	(46)	-----	
26695A.cod	(1279)	-----	
1182B	(1429)	GCGAGAAAGTTGATTAAAAAATTGGGTTTGTA	
1218B.nuc	(1426)	GCGAGAAAGTTGATTAAAAAATTGGGTTTGTA	
ORF19C2B	(832)	-----	
Consensus	(1451)		

FIGURE 14

Oligo Structures

Lacto-N-neo-Tetraose (LNnT)

$\text{Gal}\beta 1-4\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$

Lacto-N-Fucopentaose III (LNFP III)

$\text{Gal}\beta 1-4\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$

3

|

1 α Fucose

Linkage Analysis by GC/MS

FIGURE 15

The samples were methylated, hydrolyzed, reduced with sodium borodeuteride, acetylated and analyzed by GC/MS along with samples of LNT and LNF3.

- ▶ A Glc vs. Glc-NAc value close to 1 favors fucosylation of Glc-NAc.
- ▶ A Glc vs. Glc-NAc value close to 0 favors fucosylation of Glc

<i>H. Pylori</i> Strain	Glc vs. Glc-NAc
915A2	0.982
19C2A5	0.040
1111A2	0.975
19C2B1	0.991
1182B3	0.983

FIGURE 16

1 Liter LNFIII Synthesis

Batch Number	Resin Type	Total Yield	Actual Percent Recovery
1-02	MR3 NH_4HCO_3 column (1ml resin/1ml synthesis)	1.567 g	61%
2-02	MR3 NH_4HCO_3 column (1ml resin/1ml synthesis)	1.760 g	68%
3-02	Dowex1/Dowex 50 (2ml resin/1ml synthesis)	1.221 g	47%